BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 537 626 A1**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **92117191.4**

(51) Int. Cl.5: **G01N 33/576**, C07K 15/00

(22) Date of filing: **08.10.92**

(30) Priority: **08.10.91 JP 260824/91**

(43) Date of publication of application:
21.04.93 Bulletin 93/16

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB IT LI NL PT SE

(71) Applicant: **NATIONAL INSTITUTE OF HEALTH**
10-35, Kamioosaki 2-chome
Shinagawa-ku Tokyo(JP)

(72) Inventor: **Miyamura, Tatsuo**
4-21-22-113, Hamadayama
Suginami-ku, Tokyo(JP)
Inventor: **Saito, Izumu**
2-37-15-412, Yoyogi
Shibuya-ku, Tokyo(JP)
Inventor: **Harada, Shizuko**
1-1-14, Yutaka-machi
Shinagawa-ku, Tokyo(JP)
Inventor: **Honda, Yoshikazu**
2-3-16, Nishimikado
Kamakura-shi, Kanagawa(JP)

(74) Representative: **Kolb, Helga, Dr. Dipl.-Chem. et al**
Hoffmann, Eitle & Partner Patentanwälte
Arabellastrasse 4
W-8000 München 81 (DE)

(54) Diagnostic reagent for hepatitis C.

(57) A diagnostic reagent for hepatitis C, which detects an antibody induced by infection of hepatitis C virus, characterized in that said diagnostic reagent comprises the second envelope protein or the first non-structural protein which is encoded by the gene of hepatitis C virus and has a suger chain. This invention also provide a method for detecting an anti-hepatitis C virus antibody. The use of the diagnostic reagent for hepatitis C according to the present invention makes highly sensitive diagnosis of hepatitis C possible.

EP 0 537 626 A1

BACKGROUND OF THE INVENTION

This invention relates to a diagnostic reagent for hepatitis C comprising an antigen protein translated from a genome of hepatitis C virus. More specifically, this invention relates to a diagnostic reagent for detecting an antibody against hepatitis C virus (hereinafter referred to as "HCV"), which comprises a protein encoded by a gene of HCV, wherein said protein is identified as a glycoprotein called the second envelope protein or the first non-structural protein (hereinafter referred to as "E2/NS1").

The first successful cloning of human hepatitis virus which had been called non-A, non-B hepatits virus was accomplished in 1988 by Chiron Co., Ltd. U.S.A and the hepatitis virus was designated HCV. Further, Chiron Co., Ltd. succeeded in expressing in a yeast a fused protein which comprises at the C-terminal the polypeptide corresponding to the region having 363 amino acid residues from the third nonstructural protein (NS3) to the forth non-structural protein (NS4) both of which are portions of nonstructural proteins of HCV and at the N-terminal human superoxide dismutase(European unexamined patent publication No. 318216) and, using this recombinant antigen, developed a diagnostic reagent for hepatitis C (Science, 244, 359-362, 362-364, (1989)).

In Japan, the Japanese Red Cross Society has been using the diagnostic reagent in the screening of blood provided by donors, which is known as "C100-3 antibody test", in order to avoid post-transfusion hepatitis since the end of 1989. However, since not all samples are effectively screened only by C100-3 antibody test, post-transfusion hepatitis is not completely avoided.

Subsequently, further investigation of HCV genomes derived from the serum of a Japanese patient by the cloning technique revealed that HCV prevailed in Japan is similar to HCV obtained by Chiron Co., Ltd. but a different strain (Protein, Nucleic acid and Enzyme,36, 1679-1691, (1991)). In addition, the use of the core protein (C) region of the structural protein, the third non-structural protein (NS3) region, the fifth non-structural protein region and the like have been proposed as more effective diagnostic reagents than C100-3 (Lancet, 337, 317-319, 1991 and Japanese unexamined patent publication (hereinafter referred to as "J. P. KOKAI") No. Hei 3-103180).

The C100-3 antibody test system has a disadvantage that the detection rate and the sensitivity are low as mentioned above. Although proteins derived from C, NS3 and NS5 regions have been proposed as more effective antigens for detection than C100-3, any satisfactory results have not yet been reported. Therefore, there is a need for a diagnostic reagent and a diagnostic method for hepatitis C, having a higher detection rate and sensitivity.

SUMMARY OF THE INVENTION

The inventors have conducted various investigations to obtain a diagnostic reagent for hepatitis C, having a higher detection rate and sensitivity. As a result, they have found that E2/NS1 protein having a suger chain, which is obtained by expressing cDNA of E2/NS1 region in animal cells reacts with the serum of the patient of hepatitis C with a high rate in a fluorescent antibody test and accomplished the goals of the present invention. The high reaction rate of E2/NS1 region with the serum of the patient of hepatitis C was unexpected because the protein derived from E2/NS1 region is susceptible to the mutation of an amino acid sequence and, therefore, the protein expressed in E.coli has been considered to react with the serum of the patient of hepatitis C with a lower rate comparing with the proteins derived from the other regions of HCV and it has not been expected to use the protein for a diagnostic reagent.

The present invention provides a diagnostic reagent for hepatitis C, which detects an antibody induced by infection of hepatitis C virus, characterised in that said diagnostic reagent comprises the second envelope protein or the first non-structural protein which is encoded by the genome of hepatitis C virus and has a suger chain.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the steps of constructing DNA fragment 1325SK containing the base sequence of clone J1-1325.

Fig. 2 shows the steps of constructing plasmid pSR316EP.

Fig. 3 shows the steps of constructing plasmid pSRNot.

Fig. 4 shows the steps of constructing expression vector paSR1325X-3 having a DNA fragment coding for E2/NS1 protein.

Fig. 5 shows th steps of constructing plasmid pHLp1.

Fig. 6 shows the steps of constructing expression vector mulcos pHL16SR1325 having 16 DNA fragments coding for E2/NS1 protein.

DETAILED EXPLANATION OF THE INVENTION

E2/NS1 protein of the present invention is a protein derived from the region called the second envelope protein or the first nonstructural protein, which is encoded by the genome of HCV. Examples of the proteins are illustrated in SEQUENCE ID Nos.1-12 in SUQUENCE LISTING. Proteins obtained from such proteins by deleting, inserting, modifying or adding a part of amino acids are encompassed in the scope of the present invention provided that they maintain the reactivity with the serum of the patient of hepatitis C.

(1) Method of preparing clones of cDNA derived from the serum of the patient of hepatitis C, which are shown in SEQUENCE ID Nos. 1-3 of SEQUENCE LISTING and determining the base sequence thereof

Genes or DNA fragments coding for novel polypeptides, which are shown in SEQUENCE ID Nos. 1-3 of SEQUENCE LISTING can be prepared, for example, by a method described below.

Since there exists a trace of HCV in the serum and the genome of HCV is expected to be RNA, it was expected that cloning by Okayama-Berg method or Gubler-Hoffman method of the prior art would be attended by difficulties and, therefore, the following method was conducted to ensure the cloning of the gene susceptible to mutation from a trace of the serum.

The nucleic acid is extracted from the serum of the patient of hepatitis C as described in Example 1 later. Generally, it is preferred to use the serum having an OD value of 3.5 or more measured by a test kit of Ortho Inc. However, the present invention is not limited to the use of the serum having such an OD value. The serum is preferably mixed with transfer RNA (tRNA) as a carrier of virus RNA. The carrier is not limited to tRNA. Any polyribonucleoside can be used as carriers. If tRNA is used, there is an advantage that it can be rapidly confirmed by electrophoresis whether there is a required amount of tRNA having an intact length. By this confimation, it can also be confirmed whether virus RNA degradates after being mixed with tRNA as a carrier of virus RNA. As a technique of cloning cDNA from the nucleic acid, it is preferred to use polymerase chain reaction method developed by Saiki et al. (PCR method, Nature, 324, 126, (1986)). First of all, a reverse transcriptase is reacted using virus RNA as a template. In the reaction, any commercially available random primers or synthesized DNA having a base sequence similar to that of primer AS1 which is shown below may be used as a primer.

5' 3'

AS1:GCTATCAGCAGCATCATCCA SEQUENCE ID No.13

A few bases at the 5' end of these sequences may be changed to other bases. Preferably, a few bases within 10 bases from the 5' end and more preferably, a few bases within 5 bases from the 5' end may be changed to other bases. In addition, 4-5 bases, preferably a few bases may be deleted from the sequences at the 5' end of these sequences. Furthermore, any 8-12 bases, preferably 5-6 bases, more preferably a few bases, may be added to the sequences at the 5' end of these sequences.

PCR method is specifically carried out under the conditions described in Example 1. PCR method is carried out as described in Example 1 using the first complementary DNA (1st cDNA) thus obtained as a template to prepare a desired DNA fragment. The conditions of PCR method are suitably selected depending on the cicumstances. Representative examples of sense primers include the following one:

5' 3'

S1:CAGITAITCCGGATCCCICAAG SEQUENCE ID No.14

"I" appearing in the sequenc means inosine. A few bases at the 5' end of these sequences may be changed to other bas s. Preferably, a few bases within 10 bases, more preferably, within 5 bases from the 5' end may b changed to oth r bases. In addition, 4-5 bases, preferably a few bases may be deleted from the sequences at the 5' end of these sequences. Furthermore, any 8-12 bases, preferably 5-6 bases, more

preferably a few bases may b added to the sequences at the 5' nd of these sequenc s.

The DNA fragment thus obtained is inserted at one of cloning sites such as Sma I site of a cloning vector such as pUC19 according to conventional technique. Using a plasmid having this DNA fragment, the base sequences of at least 3 clones are determined independently regarding the both strands. The determination of the base sequences can be easily carried out by a dideoxy method using, for example, 7-deaza sequence kit availble from Takara Shuzo Co.,Ltd. or fluorescence sequencer GENESIS 2000 system available from Du Pont according to the protocol thereof. When the DNA fragment has a site which is considered difficult to determine the base sequence or has more than about 180 base pairs, a subcloning may be carried out according to conventional technique. SEQUENCE ID Nos.1-3 of SEQUENCE LISTING show the amino acid sequences of the proteins assumed from the base sequences of the DNA fragments thus determined.

Clone J1-1325 (SEQUENCE ID No.1), clone N27, clone N19, H19 and Y19 (SEQUENCE ID No. 3) were prepared with the serums of different patients. Clone MX24 (SEQUENCE ID No.3) was prepared with a pool of the serums of the patients of hepatitis C. The clones shown in SEQUENCE ID Nos.1-3, which were prepared using a combination of primer S1 with primer AS1 correspond to the same region in the gene of HCV.

Antigen proteins derived from E2/NS1 protein regions shown in SEQUENCE ID Nos.4-12 of SEQUENCE LISTING can also be used in the present invention.

The antigen protein of SEQUENCE ID No.4 can be obtained by expressing cDNA described in Journal of Virology, 65, 1105-1113, (1991). The antigen protein of SEQUENCE ID No.5 can be obtained by expressing cDNA described in Proceedings of the National Academy of Sciences of the USA, 87, 9524-9528, (1990). The antigen protein of SEQUENCE ID No.6 can be obtained by expressing cDNA described in The fiftieth general meeting of Japanese Cancer Society, 379, (1991). The antigen protein of SEQUENCE ID No.7 can be obtained by expressing cDNA described in European Patent No.0,388,232 (1990). The antigen proteins of SEQUENCE ID Nos.8 and 9 can be obtained by expressing cDNAs described in Proceedings of the National Academy of Sciences of the USA, 88, 3392-3396, (1991). The antigen proteins of SEQUENCE ID Nos.10 and 11 can be obtained by expressing cDNAs described in Japanese Journal of Experimental Medicine, 60, 167-177, (1990). The antigen protein of SEQUENCE ID No.12 can be obtained by expressing cDNA described in Biochemical and Biophysical Research Communications, 175, 220-228, (1991). The sequences shown in SEQUENCE ID Nos.4-12 correspond to the same region as that of the sequences shown in SEQUENCE ID Nos.1-3. (2) Expression of polypeptides encoded by the clones prepared in step (1)

In order to produce E2/NS1 protein, it is necessary to select an appropriate host-vector system which is able to stably express the protein. Further, it is required that the expressed E2/NS1 protein has the same level of biological activity, that is, antigenicity as that of HCV. Considering that natural E2/NS1 protein is expected to be a glycoprotein and that E2/NS1 protein contains many cysteine residues and the positions of the thiol bonds between the cysteine residues and the higher-order structure of the protein are important to maintain the activity, it is desired to express the protein in such an animal cell host as CHO cell, COS cell, mouse L cell, mouse C127 cell and mouse FM3A cell, preferably CHO cell. When these cells are used as hosts, it is expected that processed E2/NS1 protein is produced by introducing E2/NS1 gene having a signal-like sequence of from the 32 position to the 44 position of the amino acid sequences shown in SEQUENCE ID Nos.1-12 into the cell. Expression plasmids for these animal host cells can be constructed as follows:

As promoters in the animal cells, one can use the active-type promoter of adenovirus EIA gene (Biochemical Experiment Lecture, second series, Vol. 1, Techniques for gene investigations II, 189-190 (1986)), the early promoter of SV40, the late promoter of SV40, the promoter of apolipoprotein E gene and SR α promoter (Molecular and Celluar Biology, 8, 466-472, (1988)), preferably the promoter of SV40 and SR α promoter.

A DNA fragment of a gene coding for E2/NS1 protein containing the signal-like sequence is inserted downstream of the promoter in a direction of the transcription. When the expression vector of E2/NS1 protein is constructed, a ligated gene fragment of at least two gene fragments coding for E2/NS1 protein may be inserted downstream of the promoter. At least two units of DNA fragments ligated upstream of the 5' end of the D NA fragment of the gene coding for E2/NS1 protein with such a promoter as that of SV40 may be ligated together in the same direction of the transcription and then inserted in the vector. Polyadenylation sequence is required to be present downstream of the gene coding for E2/NS1 protein. For example, at least one of polyadenylation sequences derived from SV40 gene, β-globin gene or metallothionein gene is required to be pr s nt downstream of th g ne coding for E2/NS1 prot in. When at least two of the DNA fragments containing the gene coding for E2/NS1 protein ligated to the promoter are

ligated, th polyadenylation s quenc may be pr s nt at ach 3' end of the gen coding for E2/NS1 protein.

In transforming an animal cell such as CHO cell with this expression vector, the use of a selective marker is desired. Examples of the selective markers include DHFR gene expressing methotrexate resistance (Journal of Molecular Biology, 159, 601, (1982)), Neo gene expressing antibiotic G-418 resistance (Journal of Molecular Applied Genetics, 1, 327, (1982)), Ecogpt gene derived from E. coli, expressing mycophenol acid resistance (Proceedings of the National Academy of Sciences of the USA, 78, 2072, (1981)), hph gene expressing antibiotic hygromycin resistance (Molecular and Celluar Biology, 5, 410, (1985)) and the like. A promoter such as the aforementioned promoter derived from SV40 and the promoter of TK gene of Herpes virus is inserted upstream of the 5' end of each drug resistance gene. The aforementioned polyadenylation sequence are contained downstream of the 3' end of each drug resistance gene. When such a drug resistance gene is inserted in the expression vector of E2/NS1 protein, it may be inserted downstream of the polyadenylated site in the gene coding for E2/NS1 protein in a right direction or a reverse direction. These expression vectors do not require any co-transfection with another plasmid containing a selective marker gene in preparing a transfect.

In the case where such a selective marker gene is not inserted in the expression vector of E2/NS1 protein, a vector having a selective marker of the transfect, such as pSV2neo (Journal of Molecular Applied Genetics, 1, 327, (1982)), pMBG (Nature, 294, 228, (1981)), pSV2gpt (Proceedings of the National Academy of Sciences of the USA, 78, 2072, (1981)), pAd-D26-1 (Journal of Molecular Biology, 159, 601, (1982)) and the like may be used together with the expression vector of E2/NS1 protein to conduct co-transfection. The transfect can be easily selected by gene expression of the selective marker gene.

Examples of methods of introducing the expression vector into the animal cell include calcium phosphate method (Virology, 52, 456, (1973)) and electroporation method (Journal of Membrane Biology, 10, 279, (1972)). Calcium phosphate method is used in general.

The transfected animal cell can be cultured by a float culture or an adherent culture in the conventional manner. The cultivation can be conducted in a medium such as MEM, Ham, F-12 and the like in the presence of 5-10 % of serum or a suitable amount of insulin, dexamethasone and transferrin or in the absence of serum. The animal cell expressing E2/NS1 protein can be detected by fluorescent antibody technique using the serum of the patient according to the conventional method. The cloning is carried out by limiting dilution according to the conventional method to establish a cell line stably producing E2/NS1 protein.

E2/NS1 protein derived from HCV gene, thus obtained can be used as HCV antigen which reacts immunologically with the serum containing HCV antibody and therefore, is useful for the confirmation or the detection of the presence of Anti-HCV antibody in samples including blood or serum. Examples of the immunoassays include RIA (radioimmunoassay), ELISA (enzyme-linked immunoadosorbent assay), fluorescent antibody technique, agglutination reaction including latex fixation, immuno precipitation and the like. In the detection, a labelled antibody is usually used. A labelling substance such as a fluorescent substance, a chemoluminescent substance, a radioactive substance, a dyeing substance and the like can be used. Accordingly, using the above E2/NS1 protein derived from HCV gene as an antigen, the diagnostic reagent for hepatitis C according to the present invention can be prepared.

The reagent containing the protein having a sugar chain, which is derived from E2/NS1 region according to the present invention makes the confirmation or the detection of the presence of anti-HCV antibody in samples including blood or serum possible. The use of the reagent according to the present invention makes highly sensitive diagnosis of hepatitis C possible.

The present invention will be explained in more detail with reference to the following non-limiting examples.

Example 1

(1) Extraction of the nucleic acid from the serum of the patient of hepatitis C

Twenty-five milliliters of a Tris buffer (50 mM Tris-HCl (pH 8.0), 1 mM EDTA and 100 mM NaCl) were added to 10 ml of the serum of the patient of hepatitis C, which showed at least 3.5 of an OD value by a HCV EIA kit available from Ortho Inc. After being mixed, the mixture was centrifuged at 20,000 x g at 20 °C for 20 minutes. The obtained supernatant was centrifuged at 100,000 x g at 20 °C for additional 5 hours. One point fiv milliliters of a Prot nase K solution (1% sodium dod cyl sulfate, 10 mM EDTA, 10mM Tris-HCl (pH 7.5), 2 mg/ml Protenase K (availabl from Pharmacia Co.) and 6.6 μ g of a yeast tRNA mixtur) were added to the precipitate. After the precipitate was dissolved in the Protenase K solution, the obtained

solution was maintained at 45 °C for 90 minutes. Th mixture was subjected at least four tim s to a phenol/chloroform treatment which comprises the steps of adding an equivalent amount of phenol/chloroform, violently agitating and then c ntrifuging the mixture to coll ct an aqueous phase containing a nucleic acid. Then, a chloroform treatment was carried out at least 2 times. To the obtained aqueous phase, one-tenth amount of 3M sodium acetate or an equivalent amount of 4M ammonium acetate, and 2.5-fold volume of ethanol were added and the mixture was left to stand at -20 °C overnight or -80 °C for at least 15 minutes. The mixture was centrifuged at 35,000 rpm for 4 hours by a SW41Ti rotor (available from Beckmann Co.) to collect a nucleic acid as a precipitate.

(2) Synthesis of cDNA

(2-1) Synthesis of an RNA sample

After the nucleic acid obtained in step (1) was dried, 30 μl of water and 10 μl of ribonuclease inhibitor (100 units/ μl, available from Takara Shuzo Co., Ltd.) were added thereto to dissolve the nucleic acid. The following synthesis of cDNA was carried out using the obtained nucleic acid solution.

(2-2) Synthesis of cDNA using an anti-sense primer

To 2 μl of the aqueous solution of the nucleic acid prepared in step (2-1), 1 μl of an anti-sense primer (synthesized DNA primer AS1 ; 15 pmoles/ μl), 2 μl of 10xRT buffer (100mM Tris-HCl (pH 8.3) and 500 mM of KCl), 4 μl of 25 mM $MgCl_2$, 8 μl of 2.5 mM 4dNTP and 1 μl of water were added and the mixture was maintained at 65 °C for 5 minutes and at room temperature for 5 minutes. Subsequently, 1 μl of 25 units of a reverse transcriptase (available from Life Science Co.) and 1 μl of a ribonuclease inhibitor (100 units/ μl, available from Takara Shuzo Co., Ltd.) were added to the mixture and then the resulting mixture was maintained at 37 °C for 20 minutes, then at 42 °C for 30 minutes and finally at 95 °C for 2 minutes. Immediately thereafter, the mixture was cooled to 0 °C (Synthesis of complementary DNA). The DNA having a specific sequence was amplified using 10 μl of the DNA sample according to Saiki's method (Nature, 324, 126, (1986)), so-called PCR method as follows:

Water was added to a mixture of 10 μl of the above DNA sample, 10 μl of 10xPCR buffer (100 mM of Tris-HCl (pH 8.3), 500 mM of KCl, 15 mM of $MgCl_2$, and 1 % of gelatin), 8 μl of 2.5 mM 4dNTP, 2 μl of the synthesized DNA primer used in the synthesis of the complementary DNA (150 pmoles/ μl), 3 μl of a synthesized DNA primer corresponding to the DNA primer (15 pmoles/ μl) (which is complementary to the synthesized DNA primer used in the synthesis of the complementary DNA, i.e., the aforementioned primer S1) to prepare 100 μl of an aqueous solution. After the solution was maintained at 95 °C for 5 minutes, it was cooled rapidly to 0 °C. One minute after the cooling, the solution was mixed with 0.5 μl of Taq DNA polymerase (7 units/ μl, Trade Name "AmpliTaqTM" available from Takara Shuzo Co., Ltd.) and then mineral oil was layered on the mixture. This sample was incubated on a DNA Thermal Cycler available from Parkin Elmer Cetus Co. at 95 °C for 1 minute, at 40-55 °C for 1 minute, and at 72 °C for 1-5 minutes for 25 cycles. After the sample was incubated finally at 72 °C for 7 minutes, the reaction aqueous solution was subjected to a phenol/chloroform treatment and a precipitation treatment with ethanol to obtain amplified DNA fragments. The above precipitation treatment with ethanol was carried out by mixing the aqueous phase with a one-tenth amount of 3 M sodium acetate or an equivalent amount of 4 M ammonium acetate together with a 2.5-fold volume of ethanol, centrifuging the mixture at 15,000 rpm at 4 °C for 15 minutes by a rotor having a radius of about 5 cm and drying the precipitate.

(3) Cloning of the amplified DNA fragments and Determination of the base sequences thereof

At least 1 pmole of the DNA fragments obtained by the method described in step (2-2) was treated with T4 DNA polymerase (available from TOYOBO CO.,LTD) to make blunt ends (Molecular Cloning, 1982, Cold Spring Harbor Laboratory Press). After a phosphoric acid group was introduced into the DNA fragment at the 5' end with polynucleotidekinase (available from TOYOBO CO.,LTD) (Molecular Cloning, 1982, Cold Spring Harbor Laboratory Press), the DNA fragment was inserted at Sma I site present in the multicloning sites of pUC19 cloning vector using a ligation kit (available from Takara Shuzo Co., Ltd.).

The vector DNA prepared in the following procedur was used in the ligation in an amount of 5-10 ng.

pUC18 cloning vector was cleaved with restriction enzyme Sma I (available from TOYOBO CO.,LTD) and then subjected to a phenol/chloroform treatment and a precipitaion treatm nt with ethanol. Subsequently, this was treated with alkaline phosphatase (available from Boehringer Mannheim) to conduct the

dephosphorylation at th 5' end (Molecular Cloning, 1982, Cold Spring Harbor Laboratory Press), followed by a phenol/chloroform treatment and a precipitation with ethanol. The competent c ll of E.coli JM109 or DH5 (available from TOYOBO CO.,LTD) was transformed with the DNA prepared in the above procedure. The procedure of the transformation was according to the protocol of COMPETENT HIGH prepared by TOYOBO CO.,LTD. At least 20 transformants transformed with the pUC18 cloning vector having the DNA fragment obtained by the method described in step (2-2) using the combination of the aforementioned primers were prepared.

Plasmid DNA pUC1325 shown in Fig. 1 was prepared from the obtained transformant in the conventional method and the base sequence of the plasmid was determined by a 7-deaza sequence kit available from Takara Shuzo Co., Ltd. or a fluorescence sequencer GENESIS 2000 system available from Du Pont. Two kinds of synthesized primers, 5'd(GTAAAACGACGGCCAGT)3' (SEQUENCE ID No. 15) and 5'd-(CAGGAAACAGCTATGAC) 3' (SEQUENCE ID No. 16) were used to determine a base sequence of the + strand and that of the - strand of the DNA fragment. The DNA fragment had the same base sequence as that shown in SEQUENCE ID No. 1 of SEQUENCE LISTING. The amino acid sequence shown in SEQUENCE ID No. 1 of SEQUENCE LISTING is encoded by the + strand of the gene derived from HCV and inserted in the plasmid of the transformant.

The amino acid sequence encoded by the DNA fragment obtained was compared with the reported sequences of hepatitis C viruses. In step (2-2) of Example 1, three clones were obtained from the serum of one patient. The determination of the base sequence of the clones reveals that the patient carries several kinds of viruses.

(4) Preparation of a plasmid expressing E2/NS1 protein

Figs. 1-6 show a procedure of preparing a plasmid expressing E2/NS1 protein.

(4-1) Preparation of DNA fragment 1325SK

The DNA fragment of clone 1325 contained in plasmid pUC1325 obtained in step (3) was inserted at Sma I site of pUC18 so that the fragment had KpnI site of pUC18 at the 5' end of the + strand of clone 1325 coding for E2/NS1 protein and Bam HI site of pUC18 at the 3' end. After complete digestion with restriction enzyme Hin dIII, the fragment was partially digested with restriction enzyme Bam HI to obtain a DNA fragment which was cleaved not at Bam HI site within the vector but only at another Bam HI site present in clone 1325. The DNA fragment contains from the Bam HI site present at the 5' end to the 3' end of clone 1325 which was the DNA fragment obtained in step (2-2), which was derived from the gene of HCV.

Subsequently, as shown in Fig. 1, the DNA fragment was treated with T4 DNA polymerase to make blunt ends. After being ligated with SpeI linker consisting of the sequence of 5' pGGACTAGTCC 3' (SEQUENCE ID No. 17) (available from New England Biolab Co.), the fragment was cleaved with restriction enzyme Xba I (the Xba I site of the fragment was derived from plasmid pUC18). The following adaptor was ligated to Xba I site at the 3' end to obtain DNA fragment 1325SK.

```
5' pCTAGAGAATTCGGTAC 3' (SEQUENCE ID No. 18)
3'       TCTTAAGCp    5'
```

(4-2) Construction of plasmid pSRNot

Expression vector pAC316 reported in Journal of Virology, 65, 3015-3021, (1991) was cleaved with restriction enzyme Tth 111I at Tth111I site present at the 3' end of 3' poly A region. T4 DNA polymerase was acted on the cleaved vector to make blunt ends. The fragment between SalI sit and Eco RI site of plasmid pmoRH (Fig. 2) reported by Ikeda et al (Gene, 71, 19-27, (1988)) was cut out and T4 DNA polymerase was acted on the fragment to mak blunt ends.

As shown in Fig. 2, the DNA fragment derived from pAC316 and the DNA fragm nt derived from pmoRH were ligated together with Bgl II link r (available from Takara Shuzo Co., Ltd.) to obtain plasmid pSR316EP containing one BglII linker and one DNA fragment containing the early promoter of SV40 derived

from pmoRH. As shown in Fig. 3, after plasmid pSR316EP was cleaved with restriction nzymes Hgi AI and Dra III, T4 DNA polymerase was acted on th plasmid to make blunt ends. Then, one Not I linker was introduced in the plasmid to obtain plasmid psRNot (Fig. 3). Namely, NotI linker was pr pared by synthesizing DNA having a sequence of 5' AGCGGCCGC 3' and phosphorylating the 5' end by kination (Molecular Cloning second eddition, 11.31-11.44, (1989), Cold Spring Harbor Labratory Press).

Subsequently, dhfr gene was cut out from plasmid pCHD2L reported by Ikeda et al in Gene, 71, 19-27, (1988) using restriction enzymes Kpn I and Eco RV and Kpn I- EcoRV fragment of plasmid Charomid9-36 described in Proceedings of the National Academy of Sciences of the USA, 83, 8664-8668, (1986) was inserted in the deleted dhfr gene region instead of the KpnI- EcoRV fragment coding for dhfr gene as shown in Fig. 5 to obtain plasmid pChmBp1. The plasmid contains a polylinker derived from plasmid Charomid9-36.

Then, plasmid pAG60 reported by Garapin et al. in Journal of Molecular Biology, 150 , 1-14, (1981) was cleaved with restriction enzyme Pvu II to obtain a Pvu II fragment coding for a neomycin gene. After plasmid pChmBp1 was cleaved with restriction enzyme Eco RV and then T4 DNA polymerase was acted to make blunt ends, the fragment obtained was ligated to the Pvu II fragment to obtain plasmid pHLp1 which contained the neomycin gene derived from plasmid pAG60 at the Eco RV site of plasmid pChmBp1 (Fig. 5).

(4-3) Construction of expression vector paSR1325X-3

As shown in Fig. 4, after plasmid pSRNot obtained in step (4-2) was cleaved with restriction enzyme Not I and then with T4 DNA polymerase to make blunt ends, this was cleaved with restriction enzyme Kpn I. The obtained DNA fragment was ligated to DNA fragment 1325SK obtained in step (4-1) to obtain expression vector paSR1325X-3 having only one DNA fragment 1325SK (Fig. 4).

(4-4) Construction of expression vector pHL16SR1325

As shown in Fig. 6, expression vector paSR1325X-3 obtained in step (4-3) was cleaved with restriction enzyme Sfi I to prepare two fragments one of which was an expression unit of clone 1325. The Sfi I sites were present in an initial promoter of SV40. Five μg of the Sfi I fragment having the expressin unit of clone 1325 was ligated to 50 ng of the fragment obtained by cleaving expression vector pHLp1 with restriction enzyme Sfi I in 10 μl of a reaction solution using a ligation kit available from Takara Shuzo Co., Ltd. according to a protocol for the ligation kit to obtain expression vector pHL16SR1325 (Fig. 6).

The vector had successive sixteen DNA fragments 1325SK having at the Sfi I site of expression vector paSR1325X-3 the expression unit of clone 1325 which was a gene coding for E2/NS1 protein. In the vector, all of the DNA fragments 1325SK were inserted downstream of SV40 promoter of expression vector paSR1325X-3 in a direction of transcription.

(5) Obtaining a cell line constantly expressing E2/NS1 protein

Expression vector pHL16SR1325 prepared in step (4) was recovered from the recombinant E.coli DH1 strain, purified according to the conventional technique described in Molecular Cloning second edition, 1989, Cold Spring Harbor Laboratory Press to obtain a large amount of the expression plasmid DNA. CHO cells were transfected with the plasmid DNA according to the method described in Ausubel et al. (Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience, Capter 9.1.1-9.1.4, (1987)) as follows:

CHO cells were cultured in Ham F-12 medium containing 10 % of fetal calf serum (FCS) in a plate having a diamer of 6 cm until the cells were in semiconfluent condition. Then, the medium was removed from the plate and a DNA solution was dropwise added thereto. The DNA solution was previously prepared by the following procedure.

Three hundreds μl of 2xHEBS solution (2xHEBS solution ; 1.6 % sodium chloride, 0.074 % potassium chloride, 0.05 % $Na_2HPO_4 \cdot 12H_2O$, 0.2 % dextrose and 1 % HEPES (pH 7.05)) were mixed with 10 μg of the plasmid DNA in each plate and steriliz d wat r was added to the mixture to prepare a solution of 570 μl. The solution was charged in an Eppendorf centrifuge tube. The DNA solution was violently agitated by a Vortex mixer for 1-2 seconds while adding 30 μl of 2.5 M calcium chlorid solution th reto. The DNA solution was agitated by a Vortex mixer at about 10-minute intervals during being left to stand for 30 minutes. Th obtained DNA solution was add d to the afor m ntioned CHO c lls and the CHO cells were left to stand at room temperature for 30 minutes. Then, 5 ml of Ham F-12 medium containing 10 % of FCS

available from GIBCO Co. were add d to the plate and the cultur was incubated at 37 °C under air containing 5 % carbon dioxide for 4-5 hours. Subsequently, the medium was removed from the plate and the cells wer washed with 5 ml of a 1xTBS + + solution (1xTBS + + solution ; 25 mM Tris-HCl (pH 7.5), 140 mM sodium chloride, 5mM potassium chloride, 0.6 mM disodium hydrogen phosphate, 0.08 mM calcium chloride and 0.08 mM magnesium chloride). After the 1xTBS + + solution was removed, 5 ml of a 1xTBS + + solution containing 20 % of glycerol was added to the cells and the culture was left to stand at room temperature for 1-2 minutes. After the supernatant was removed from the plate, the cells were washed again with 5 ml of a 1xTBS + + solution and cultured in 5 ml of fresh Ham F-12 medium containing 10 % of FCS in the plate at 37 °C under air containing 5 % carbon dioxide for 48 hours. Then, the medium was removed and the cells were washed with 5 ml of a 1xTBS + + solution. The cells were treated with a trypsin-EDTA solution (available from Sigma Co.) and left to stand at room temperature for 30 seconds. Five minutes after the trypsin-EDTA solution was removed, the cells attached to the wall of the plate were peeled adding 5 ml of Ham F-12 medium containing 10 % of FCS. The cells cultured in one plate having a diameter of 5 cm were divided in ten plates having a diameter of 9 cm and cultured in the plates containing drug G418 (G418 sulfate (GENETICIN) available from GIBCO Co.) in a concentration of 600 μ g /ml. Ten days after the cultivation, grown cells having G418 resistance were isolated and cultured for about 7 days in 1 ml of Ham F-12 medium containing 10 % of FCS in a 24 well titer plate each well of which has an area of about 3.1 cm^2.

A part of the cells were clutured on slide glass (Lab-Tek Chamber Slides, Nunc4808 available from Japan Inter Med Co.) overnight. After being rinsed with phosphate buffered saline (PBS), the slide glass was immersed in cold actone-methanol (1:1) solution and maintained at -20 °C for 15 minutes to fix the cells. The cells fixed on the slide glass were reacted with the serum of the patient of hepatitis C 20-fold diluted with PBS at 37 °C for 30 minutes. Then, the slide glass was washed three times with PBS for 5 minutes and reacted with FITC-labelled rabbit anti-human IgG (available from Daco Japan Co.) 50-fold diluted with PBS at 37 °C for 30 minutes. The slide glass was washed three times with PBS for 5 minutes and dried by putting the slide glass between two pieces of filter paper. After the slide glass was sealed with glycerin, the cells on the slide glass were observed under a fluorescence microscope. Screening positive cells as descrived above, successive three times of limiting dilution were carried out to establish cell line 13L20 constantly producing E2/NS1 protein.

(6) Study of the reactivity of 13L20 cells with the serum of the patient of hepatitis C

After 13L20 cells established in step (5) were cultured on Lab-Tek Chamber Slides (Lab-Tek Chamber Slides, Nunc4808 available from Japan Inter Med Co.) overnight and then fixed with a cold acetone-methanol solution, the fixed cells were reacted with 59 serum samples of the patients of hepatitis C. Then, the cells were washed as described above and reacted with the secondary antibody. The observation under a fluorescence microscope revealed that 53 samples were positive. Among the 59 serum samples, 6 samples were judged to be positive using CHO cells constantly producing the first envelope region of HCV.

Example 2

Using as a template the DNA fragment described in Example 11 (3) of the specification of European Patent Application No. 92109812.5 filed on June 11, 1992 (TITLE OF THE INVENTION "Gene or DNA fragments derived from hepatitis C virus, polypeptides encoded thereby, and method of producing thereof"), PCR reaction was carried out in the same manner as that of Example 1 using the same primer to obtain a DNA fragment corresponding the same region as that of clone J1-1325 shown in SEQUENCE ID No. 1 of SEQUENCE LISTING. The region was a DNA fragment encoding for E2/NS1 protein like clone J1-1325. For example, using as a template the DNA fragment clone N27MX24A-1 having a base sequence shown in SEQUENCE ID No. 31 of SEQUENCE LISTING described in the specification of the aforementioned European Patent Application filed on June 11, 1992, plasmid pUCN27MX24A-2 was obtained. The base sequence of the DNA fragment coding for E2/NS1 protein, which was cloned in the plasmid is shown in SEQUENCE ID No. 2 of SEQUENCE LISTING. In addition, MK2724A2 cell line constantly producing E2/NS1 protein was stablised by the same procedur as that described in steps (4) and (5) of Example 1. The reactivity of th same samples as Example 1 with the cell line was estimated by the same method as that described in step (6) of Example 1. Results similar to those obtained in step (6) of Example 1 were obtained.

SEQUENCE LISTING

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: J1-1325

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
G ATC CCA CAA GCT GTC ATG GAC ATG GTG GCG GGG GCC CAC TGG GGA GTC     49
 Ile Pro Gln Ala Val Met Asp Met Val Ala Gly Ala His Trp Gly Val
   1               5                  10                  15

CTA GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT      97
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
             20                  25                  30

TTG ATT GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG CAT ACC CGC GTG     145
Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly His Thr Arg Val
         35                  40                  45

ACG GGG GGG GTG CAA GGC CAT GTC ACC TCT ACA CTC ACG TCC CTC TTT     193
Thr Gly Gly Val Gln Gly His Val Thr Ser Thr Leu Thr Ser Leu Phe
     50                  55                  60
```

```
AGA CCT GGG GCG TCC CAG AAA ATT CAG CTT GTA AAC ACC AAT GGC AGT     241
Arg Pro Gly Ala Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly Ser
 65                  70                  75                  80
TGG CAT ATC AAC AGG ACT GCC CTG AAC TGC AAT GAC TCC CTC AAA ACT     289
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Lys Thr
                85                  90                  95
GGG TTT CTT GCC GCG CTG TTC TAC ACA CAC AAG TTC AAC GCG TCC GGA     337
Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Lys Phe Asn Ala Ser Gly
            100                 105                 110
TGC CCG GAG CGC ATG GCC AGC TGT CGC TCC ATT GAC AAG TTC GAC CAG     385
Cys Pro Glu Arg Met Ala Ser Cys Arg Ser Ile Asp Lys Phe Asp Gln
        115                 120                 125
GGA TGG GGT CCC ATC ACC TAT GCT CAA CCT GAC AAC TCG GAC CAG AGG     433
Gly Trp Gly Pro Ile Thr Tyr Ala Gln Pro Asp Asn Ser Asp Gln Arg
    130                 135                 140
CCG TAT TGC TGG CAC TAC GCA CCT CGA CAG TGT GGT ATC GTA CCC GCG     481
Pro Tyr Cys Trp His Tyr Ala Pro Arg Gln Cys Gly Ile Val Pro Ala
145                 150                 155                 160
TCG CAG GTG TGC GGT CCA GTG TAT TGC TTC ACC CCA AGC CCT GTT GTA     529
Ser Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
                165                 170                 175
GTG GGG ACG ACC GAT CGT TTC GGC GCC CCT ACG TAT AAC TGG GGG GAC     577
Val Gly Thr Thr Asp Arg Phe Gly Ala Pro Thr Tyr Asn Trp Gly Asp
            180                 185                 190
AAT GAG ACG GAC GTG CTG CTC CTA AAC AAC ACG CGG CCG CCG CAT GGC     625
Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr Arg Pro Pro His Gly
        195                 200                 205
AAC TGG TTC GGC TGT ACA TGG ATG AAT AGC ACT GGG TTC ACC AAG ACG     673
Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Thr
```

```
    210                 215                 220
TGC GGA GGC CCC CCG TGT AAC ATC AGG GGG GTC GGC AAC AAC ACC TTG     721
Cys Gly Gly Pro Pro Cys Asn Ile Arg Gly Val Gly Asn Asn Thr Leu
225                 230                 235                 240
ACC TGC CCC ACG GAC TGC TTC CGG AAG CAC CCC GAC GCC ACT TAC ACA     769
Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro Asp Ala Thr Tyr Thr
                245                 250                 255
AAA TGT GGT TCG GGC CCT TGG TTG ACA CCT AGG TGC TTG GTT GAC TAC     817
Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu Val Asp Tyr
            260                 265                 270
CCA TAC AGG CTC TGG CAC TAC CCC TGC ACT GTC AAC TTT ACC ATC TTC     865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Ile Phe
        275                 280                 285
AAG GTT AGG ATG TAT GTG GGG GGC GTG GAG CAC AGG CTT GAT GCT GCA     913
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Asp Ala Ala
    290                 295                 300
TGC AAC TGG ACT CGA GGA GAG CGT TGC GAC TTG GAG GAC AGG GAT AGA     961
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320
GCA GAG CTC AGC CCG CTA CTG CTG TCT ACG ACA GAG TGG CAG GTA CTG    1009
Ala Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp Gln Val Leu
                325                 330                 335
CCC TGT TCC TTC ACC ACC CTA CCG GCT CTG TCC ACT GGT CTA ATC CAT    1057
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350
CTC CAT CAG AAC GTC GTG GAC GTG CAA TAC CTG TAC GGT ATA GGG TCA    1105
Leu His Gln Asn Val Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser
        355                 360                 365
GCA GTT GTC TCC TTT GTA ATC AAA TGG GAG TAT GTC CTG TTG CTT TTC    1153
```

```
Ala Val Val Ser Phe Val Ile Lys Trp Glu Tyr Val Leu Leu Leu Phe
    370                 375                 380
CTT CTC CTG GCT GAC GCA CGC GTC TGT GCC TGC TTG TGG ATG ATG CTG    1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu
385                 390                 395                 400
CTG ATA                                                        1207
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: N27MX24A-2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
G ATC CCA CAA GCC GTG GTG GAT ATG GTG GCA GGG GCC CAC TGG GGA GTC     49
  Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val
    1               5                  10                  15
CTG GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTC      97
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30
TTG GTT GTG ATG CTG CTC TTC GCC GGT GTT GAC GGG GGG ACC CAC GTG     145
Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His Val
        35                  40                  45
ACA GGG GGG AAG GTA GCC TAC ACC ACC CAG GGC TTT ACA CCC TTC TTT     193
Thr Gly Gly Lys Val Ala Tyr Thr Thr Gln Gly Phe Thr Pro Phe Phe
    50                  55                  60
TCA CGA GGG CCG TCT CAG AAA ATC CAA CTT GTA AAC ACT AAC GGC AGC     241
```

```
Ser Arg Gly Pro Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly Ser
 65                  70                  75                  80
TGG CAC ATC AAT AGG ACT GCC CTC AAT TGC AAT GAC TCC CTT AAC ACC     289
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn Thr
                85                  90                  95
GGG TTC CTT GCC GCG CTG TTC TAC ACC CAC AGC TTC AAC GCG TCC GGA     337
Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Ser Phe Asn Ala Ser Gly
            100                 105                 110
TGT CCG GAG CGT ATG GCC GGT TGC CGC CCC ATT GAC GAG TTC GCT CAG     385
Cys Pro Glu Arg Met Ala Gly Cys Arg Pro Ile Asp Glu Phe Ala Gln
        115                 120                 125
GGG TGG GGT CCC ATC ACT CAT GTT GTG CCT AAC ATC TCG GAC CAG AGG     433
Gly Trp Gly Pro Ile Thr His Val Val Pro Asn Ile Ser Asp Gln Arg
    130                 135                 140
CCC TAT TGC TGG CAC TAC GCG CCT CGA CCG TGT GGT ATC GTA CCC GCG     481
Pro Tyr Cys Trp His Tyr Ala Pro Arg Pro Cys Gly Ile Val Pro Ala
145                 150                 155                 160
TCG CAG GTG TGT GGT CCG GTG TAT TGC TTC ACC CCA AGC CCT GTT GTG     529
Ser Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
                165                 170                 175
GTG GGG ACG ACC GAT CGT TTC GGC GCC CCC ACG TAC AAC TGG GGA AAC     577
Val Gly Thr Thr Asp Arg Phe Gly Ala Pro Thr Tyr Asn Trp Gly Asn
            180                 185                 190
AAT GAG ACG GAT GTG CTA CTC CTC AAC AAC ACA CGG CCG CCG CAG GGC     625
Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr Arg Pro Pro Gln Gly
        195                 200                 205
AAC TGG TTC GGT TGT ACC TGG ATG AAT GGC ACT GGG TTC ACA AAG ACG     673
Asn Trp Phe Gly Cys Thr Trp Met Asn Gly Thr Gly Phe Thr Lys Thr
    210                 215                 220
```

```
TGC GGG GGC CCC CCG TGC AAC ATC GGG GGG GTC GGC AAC AAT ACC TTG     721
Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Val Gly Asn Asn Thr Leu
225                 230                 235                 240
ACT TGC CCC ACG GAC TGC TTC CGG AAG CAC CCC GAG GCC ACT TAC ACA     769
Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Thr
                245                 250                 255
AAA TGT GGT TCG GGG CCT TGG TTG ACG CCT AGG TGC CTA GTT CAT TAC     817
Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu Val His Tyr
            260                 265                 270
CCA TAC AGG CTC TGG CAC TAT CCC TGC ACT GTC AAC TTT ACC ATC TTC     865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Ile Phe
        275                 280                 285
AAG GTT AGG ATG TAT GTG GGG GGC GTG GAA CAC AGG CTT GAA GCT GCA     913
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Glu Ala Ala
    290                 295                 300
TGC AAT TGG ACC CGA GGA GAG CGT TGT GAC TTG GAG GAC AGG GAT AGA     961
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320
TCA GAG CTT AGC CCG CTA TTG CTG TCC ACA ACA GAG TGG CAG GTA CTG    1009
Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp Gln Val Leu
                325                 330                 335
CCC TGT TCC TTC ACC ACC CTG CCG GCT CTG TCC ACT GGT TTG ATT CAT    1057
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350
CTC CAT CAG AAC ATC GTG GAC GTG CAA TAT CTG TAC GGC ATA GGG TCG    1105
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser
        355                 360                 365
GCG GTT GTC TCC TTC GCA ATC AAA TGG GAA TAT ATT CTG TTG CTT TTC    1153
Ala Val Val Ser Phe Ala Ile Lys Trp Glu Tyr Ile Leu Leu Leu Phe
```

```
    370                 375                 380
CTC CTC CTG GCG GAC GCG CGC GTC TGT GCC TGC TTG TGG ATG ATG CTG    1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu
385                 390                 395                 400
CTG ATA                                                            1207
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 402 amino acids

(B) TYPE: protein

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: N27,N19,H19,Y19,MX24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val
 1               5                  10                  15
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30
Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His Val
                                                    Arg
        35                  40                  45
Thr Gly Gly Lys Val Ala Tyr Thr Thr Gln Gly Phe Thr Pro Phe Phe
                                            Arg         Ser
```

```
                                                    Ser
     50                  55                  60
Ser Arg Gly Pro Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly Ser
                        Arg
 65                  70                  75                  80
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn Thr
                                                        Gln
                 85                  90                  95
Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Ser Phe Asn Ala Ser Gly
                Thr             Arg             Asp
            100                 105                 110
Cys Pro Glu Arg Met Ala Gly Cys Arg Pro Ile Asp Glu Phe Ala Gln
                        Ser                 Ser
        115                 120                 125
Gly Trp Gly Pro Ile Thr His Val Val Pro Asn Ile Ser Asp Gln Arg
        Asp                             Asp Val
    130                 135                 140
Pro Tyr Cys Trp His Tyr Ala Pro Arg Pro Cys Gly Ile Val Pro Ala
                                                Val
145                 150                 155                 160
Ser Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
Trp
                165                 170                 175
Val Gly Thr Thr Asp Arg Phe Gly Ala Pro Thr Tyr Asn Trp Gly Asn
                        Ser                 Thr         Ala
            180                 185                 190
Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr Arg Pro Pro Gln Gly
        195                 200                 205
Asn Trp Phe Gly Cys Thr Trp Met Asn Gly Thr Gly Phe Thr Lys Thr
```

```
    210                 215                 220
Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Val Gly Asn Asn Thr Leu
225                 230                 235                 240
Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Thr
                245                 250                 255
Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu Val His Tyr
            260                 265                 270
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Ile Phe
        275                 280                 285
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Glu Ala Ala
    290                 295                 300
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320
Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp Gln Val Leu
                325                 330                 335
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser
        355                 360                 365
Ala Val Val Ser Phe Ala Ile Lys Trp Glu Tyr Ile Leu Leu Leu Phe
    370                 375                 380
Leu Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu
385                 390                 395                 400
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: BK164

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
G ATC CCA CAA GCC GTC GTG GAC ATG GTG GCG GGG GCC CAC TGG GGA GTC     49
  Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val
    1               5                  10                  15
CTG GCG GGC CTT GCC TAC TAT TCC ATG GCG GGG AAC TGG GCT AAG GTT      97
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Ala Gly Asn Trp Ala Lys Val
            20                  25                  30
CTG ATT GTG ATG CTA CTT TTT GCT GGC GTT GAC GGG GAT ACC CAC GTG     145
Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly Asp Thr His Val
        35                  40                  45
ACA GGG GGG GCG CAA GCC AAA ACC ACC AAC AGG CTC GTG TCC ATG TTC     193
Thr Gly Gly Ala Gln Ala Lys Thr Thr Asn Arg Leu Val Ser Met Phe
    50                  55                  60
GCA AGT GGG CCG TCT CAG AAA ATC CAG CTT ATA AAC ACC AAT GGG AGT     241
Ala Ser Gly Pro Ser Gln Lys Ile Gln Leu Ile Asn Thr Asn Gly Ser
65                  70                  75                  80
TGG CAC ATC AAC AGG ACT GCC CTG AAC TGC AAT GAC TCT CTC CAG ACT     289
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Gln Thr
```

```
                    85                  90                  95
GGG TTT CTT GCC GCG CTG TTC TAC ACA CAT AGT TTC AAC TCG TCC GGG     337
Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Ser Phe Asn Ser Ser Gly
            100                 105                 110
TGC CCA GAG CGC ATG GCC CAG TGC CGC ACC ATT GAC AAG TTC GAC CAG     385
Cys Pro Glu Arg Met Ala Gln Cys Arg Thr Ile Asp Lys Phe Asp Gln
        115                 120                 125
GGA TGG GGT CCC ATT ACT TAT GCT GAG TCT AGC AGA TCA GAC CAG AGG     433
Gly Trp Gly Pro Ile Thr Tyr Ala Glu Ser Ser Arg Ser Asp Gln Arg
    130                 135                 140
CCA TAT TGC TGG CAC TAC CCA CCT CCA CAA TGT ACC ATC GTA CCT GCG     481
Pro Tyr Cys Trp His Tyr Pro Pro Pro Gln Cys Thr Ile Val Pro Ala
145                 150                 155                 160
TCG GAG GTG TGC GGC CCA GTG TAC TGC TTC ACC CCA AGC CCT GTC GTC     529
Ser Glu Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
                165                 170                 175
GTG GGG ACG ACC GAT CGT TTC GGT GTC CCT ACG TAT AGA TGG GGG GAG     577
Val Gly Thr Thr Asp Arg Phe Gly Val Pro Thr Tyr Arg Trp Gly Glu
            180                 185                 190
AAC GAG ACT GAC GTG CTG CTG CTC AAC AAC ACG CGG CCG CCG CAA GGC     625
Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr Arg Pro Pro Gln Gly
        195                 200                 205
AAC TGG TTC GGC TGC ACA TGG ATG AAT AGC ACC GGG TTC ACC AAG ACA     673
Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Thr
    210                 215                 220
TGT GGG GGG CCC CCC TGT AAC ATC GGG GGG GTC GGC AAC AAC ACC CTG     721
Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Val Gly Asn Asn Thr Leu
225                 230                 235                 240
ACC TGC CCC ACG GAC TGC TTC CGG AAG CAC CCC GAG GCT ACC TAC ACA     769
```

```
Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Thr
                245                 250                 255

AAA TGT GGT TCG GGG CCT TGG CTG ACA CCT AGG TGC ATG GTT GAC TAT   817
Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Met Val Asp Tyr
            260                 265                 270

CCA TAC AGG CTC TGG CAT TAC CCC TGC ACT GTT AAC TTT ACC ATC TTC   865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Ile Phe
        275                 280                 285

AAG GTT AGG ATG TAT GTG GGG GGG GTG GAG CAC AGG CTC AAT GCT GCA   913
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Asn Ala Ala
    290                 295                 300

TGC AAT TGG ACC CGA GGA GAG CGT TGT GAC TTG GAG GAC AGG GAT AGG   961
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320

CCG GAG CTC AGC CCG CTG CTG CTG TCT ACA ACA GAG TGG CAG GTA CTG  1009
Pro Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp Gln Val Leu
                325                 330                 335

CCC TGT TCC TTC ACC ACC CTA CCA GCT CTG TCC ACT GGC TTG ATT CAC  1057
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350

CTC CAT CAG AAC ATC GTG GAC GTG CAA TAC CTA TAC GGT ATA GGG TCA  1105
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser
        355                 360                 365

GCG GTT GTC TCC TTT GCA ATC AAA TGG GAG TAT GTC CTG TTG CTT TTC  1153
Ala Val Val Ser Phe Ala Ile Lys Trp Glu Tyr Val Leu Leu Leu Phe
    370                 375                 380

CTT CTC CTA GCG GAC GCA CGT GTC TGT GCC TGC TTG TGG ATG ATG CTG  1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu
385                 390                 395                 400
```

CTG ATA 1207

Leu Ile

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: HCV-J

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
G ATC CCA CAA GCC GTC GTG GAC ATG GTG GCG GGG GCC CAC TGG GGT GTC    49
  Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val
    1               5                  10                  15

CTA GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTC    97
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30

TTG ATT GTG ATG CTA CTC TTT GCT GGC GTT GAC GGG CAC ACC CAC GTG    145
Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly His Thr His Val
        35                  40                  45

ACA GGG GGA AGG GTA GCC TCC AGC ACC CAG AGC CTC GTG TCC TGG CTC    193
```

```
Thr Gly Gly Arg Val Ala Ser Ser Thr Gln Ser Leu Val Ser Trp Leu
     50                  55                  60
TCA CAA GGC CCA TCT CAG AAA ATC CAA CTC GTG AAC ACC AAC GGC AGC     241
Ser Gln Gly Pro Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly Ser
 65                  70                  75                  80
TGG CAC ATC AAC AGG ACC GCT CTG AAT TGC AAT GAC TCC CTC CAA ACT     289
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Gln Thr
                 85                  90                  95
GGG TTC ATT GCT GCG CTG TTC TAC GCA CAC AGG TTC AAC GCG TCC GGG     337
Gly Phe Ile Ala Ala Leu Phe Tyr Ala His Arg Phe Asn Ala Ser Gly
            100                 105                 110
TGC CCA GAG CGC ATG GCT AGC TGC CGC CCC ATC GAT GAG TTC GCT CAG     385
Cys Pro Glu Arg Met Ala Ser Cys Arg Pro Ile Asp Glu Phe Ala Gln
        115                 120                 125
GGG TGG GGT CCC ATC ACT CAT GAT ATG CCT GAG AGC TCG GAC CAG AGG     433
Gly Trp Gly Pro Ile Thr His Asp Met Pro Glu Ser Ser Asp Gln Arg
    130                 135                 140
CCA TAT TGC TGG CAC TAC GCG CCT CGA CCG TGC GGG ATC GTG CCT GCG     481
Pro Tyr Cys Trp His Tyr Ala Pro Arg Pro Cys Gly Ile Val Pro Ala
145                 150                 155                 160
TCG CAG GTG TGT GGT CCA GTG TAT TGC TTC ACT CCG AGC CCT GTT GTA     529
Ser Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
                165                 170                 175
GTG GGG ACG ACC GAT CGT TTC GGC GCT CCT ACG TAT AGC TGG GGG GAG     577
Val Gly Thr Thr Asp Arg Phe Gly Ala Pro Thr Tyr Ser Trp Gly Glu
            180                 185                 190
AAT GAG ACA GAC GTG CTG CTA CTT AGC AAC ACG CGG CCG CCT CAA GGC     625
Asn Glu Thr Asp Val Leu Leu Leu Ser Asn Thr Arg Pro Pro Gln Gly
        195                 200                 205
```

```
AAC TGG TTT GGG TGC ACG TGG ATG AAC AGC ACT GGG TTC ACC AAG ACG      673
Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Thr
    210                 215                 220

TGC GGG GGC CCT CCG TGC AAC ATC GGG GGG GTC GGC AAC AAC ACC TTG      721
Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Val Gly Asn Asn Thr Leu
225                 230                 235                 240

GTC TGC CCC ACG GAT TGC TTC CGG AAG CAC CCC GAG GCC ACT TAC ACA      769
Val Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Thr
                245                 250                 255

AAG TGT GGC TCG GGG CCC TGG TTG ACA CCC AGG TGC ATG GTT GAC TAC      817
Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Met Val Asp Tyr
            260                 265                 270

CCA TAC AGG CTC TGG CAC TAC CCC TGC ACT GTT AAC TTT ACC GTC TTT      865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Val Phe
        275                 280                 285

AAG GTC AGG ATG TAT GTG GGG GGC GTG GAG CAC AGG CTC AAT GCT GCA      913
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Asn Ala Ala
    290                 295                 300

TGC AAT TGG ACT CGA GGA GAG CGC TGT GAC TTG GAG GAC AGG GAT AGG      961
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320

TCA GAA CTC AGC CCG CTG CTG CTG TCT ACA ACA GAG TGG CAG ATA CTG     1009
Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp Gln Ile Leu
                325                 330                 335

CCC TGT TCC TTC ACC ACC CTA CCG GCC CTG TCC ACT GGC TTG ATC CAT     1057
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350

CTT CAC CGG AAC ATC GTG GAC GTG CAA TAC CTG TAC GGT ATA GGG TCG     1105
Leu His Arg Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser
```

```
        355                 360                 365
GCA GTT GTC TCC TTT GCA ATC AAA TGG GAG TAT ATC CTG TTG CTT TTC    1153
Ala Val Val Ser Phe Ala Ile Lys Trp Glu Tyr Ile Leu Leu Leu Phe
    370                 375                 380
CTT CTT CTG GCG GAC GCG CGC GTC TGT GCC TGC TTG TGG ATG ATG CTG    1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu
385                 390                 395                 400
CTG ATA                                                            1207
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: HCV-RNA33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
G ATC CCG CAA GCT GTC GTG GAC ATG GTG GCG GGG GCC CAC TGG GGA GTC    49
  Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val
    1               5                  10                  15
```

```
CTG GCG GGC CTG GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT      97
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
             20                  25                  30

TTG ATT GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG CAA ACC TAT ACG     145
Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly Gln Thr Tyr Thr
         35                  40                  45

ACG GGG GGG GCG GTT GCC CGC ACC ACC ACC GGG TTC GCG TCC CTC TTC     193
Thr Gly Gly Ala Val Ala Arg Thr Thr Thr Gly Phe Ala Ser Leu Phe
     50                  55                  60

TCC GCT GGG TCG CAG GAG AAC ATC CAG CTT ATA AAC ACC AAT GGC AGC     241
Ser Ala Gly Ser Gln Glu Asn Ile Gln Leu Ile Asn Thr Asn Gly Ser
 65                  70                  75                  80

TGG CAC ATC AAC AGG ACT GCC CTG AAC TGC AAC GAC TCC CTC AAC ACT     289
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn Thr
                 85                  90                  95

GGA TTT CTT GCC GCG CTG TTC TAC ACA CAC AAG TTC AAC TCA TCC AGA     337
Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Lys Phe Asn Ser Ser Arg
            100                 105                 110

GCC GAG AGC GTA TTG GCC AGC TGC CGC TTC ATC GAC GAG TTC GAT CAG     385
Ala Glu Ser Val Leu Ala Ser Cys Arg Phe Ile Asp Glu Phe Asp Gln
        115                 120                 125

GGA TGG GGC CCC ATC ACT TAC ACC GAG CGT AAC AGT TCG GAC CAG AGG     433
Gly Trp Gly Pro Ile Thr Tyr Thr Glu Arg Asn Ser Ser Asp Gln Arg
    130                 135                 140

CCT TAT TGC TGG CAC TAT CCA CCC CGA CAG TGT GGT ATC ATA CCC GCG     481
Pro Tyr Cys Trp His Tyr Pro Pro Arg Gln Cys Gly Ile Ile Pro Ala
145                 150                 155                 160

TCG GAG GTG TGC GGT CCA GTG TAT TGT TTC ACC CCA AGC CCT GTT GTG     529
Ser Glu Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
```

```
                    165                 170                 175
GTG GGG ACA ACC GAT CGG TTC GGT GTC CCT ACA TAC AGC TGG GGG GAG     577
Val Gly Thr Thr Asp Arg Phe Gly Val Pro Thr Tyr Ser Trp Gly Glu
                180                 185                 190
AAT GAG ACG GAC GTG CTG GTT CTC AAC AAC ACG CGG CCG CCG CAG GGC     625
Asn Glu Thr Asp Val Leu Val Leu Asn Asn Thr Arg Pro Pro Gln Gly
        195                 200                 205
AAC TGG TTC GGC TGT ACA TGG ATG AAT GGC ACT GGT TTC ACC AAG ACA     673
Asn Trp Phe Gly Cys Thr Trp Met Asn Gly Thr Gly Phe Thr Lys Thr
    210                 215                 220
TGC GGG GGT CCC CCG TGT CAC ATC GGG GGG CGC GGC AAC AAC ACC CTG     721
Cys Gly Gly Pro Pro Cys His Ile Gly Gly Arg Gly Asn Asn Thr Leu
225                 230                 235                 240
ACT TGC CCC ACG GAC TGC TTC CGG AAG CAT CCC GAG GCT ACG TAT ACA     769
Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Thr
                245                 250                 255
AAA TGT GGT TCG GGG CCT TGG TTG ACA CCT AGG TGC ATG GTT GAT TAC     817
Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Met Val Asp Tyr
            260                 265                 270
CCA TAC AGG CTC TGG CAC TAC CCC TGC ACT GTC AAC TTT ACC ACC TTT     865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Thr Phe
        275                 280                 285
AAG GTT AGG ATG TAT GTG GGG GGC GTG GAG CAC AGG CTC ATT GCT GCA     913
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Ile Ala Ala
    290                 295                 300
TGC AAT TGG ACT CGA GGA GAC CGT TGT AAC TTG GAG GAC AGG GAT AGA     961
Cys Asn Trp Thr Arg Gly Asp Arg Cys Asn Leu Glu Asp Arg Asp Arg
305                 310                 315                 320
TCA GAG CTT AGT CCG CTG CTG CTG TCT ACG ACA GAG TGG CAG ATA CTG    1009
```

```
Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp Gln Ile Leu
                325                 330                 335

CCC TGT TCC TTC ACC ACC CTA CCG GCT CTC TCC ACC GGT TTG ATC CAT    1057
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350

CTC CAT CAG AAC ATC GTG GAC GTG CAA TAC CTG TAC GGT ATA GGG TCT    1105
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser
        355                 360                 365

GCT GTT GTC TCC ATT GCA ATC AGG TGG GAA TAT GTC CTG TTG CTT TTC    1153
Ala Val Val Ser Ile Ala Ile Arg Trp Glu Tyr Val Leu Leu Leu Phe
    370                 375                 380

CTT CTC CTG GCG GAC GCG CGT GTC TGT GCC TGC TTG TGG ATG ATG CTG    1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu
385                 390                 395                 400

CTG ATA                                                            1207
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: HCV1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
G ATC CCA CAA GCC ATC TTG GAC ATG ATC GCT GGT GCT CAC TGG GGA GTC     49
  Ile Pro Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val
    1               5                  10                  15
CTG GCG GGC ATA GCG TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC      97
Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30
CTG GTA GTG CTG CTG CTA TTT GCC GGC GTC GAC GCG GAA ACC CAC GTC     145
Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala Glu Thr His Val
        35                  40                  45
ACC GGG GGA AGT GCC GGC CAC ACT GTG TCT GGA TTT GTT AGC CTC CTC     193
Thr Gly Gly Ser Ala Gly His Thr Val Ser Gly Phe Val Ser Leu Leu
    50                  55                  60
GCA CCA GGC GCC AAG CAG AAC GTC CAG CTG ATC AAC ACC AAC GGC AGT     241
Ala Pro Gly Ala Lys Gln Asn Val Gln Leu Ile Asn Thr Asn Gly Ser
65                  70                  75                  80
TGG CAC CTC AAT AGC ACG GCC CTG AAC TGC AAT GAT AGC CTC AAC ACC     289
Trp His Leu Asn Ser Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn Thr
                85                  90                  95
GGC TGG TTG GCA GGG CTT TTC TAT CAC CAC AAG TTC AAC TCT TCA GGC     337
Gly Trp Leu Ala Gly Leu Phe Tyr His His Lys Phe Asn Ser Ser Gly
            100                 105                 110
TGT CCT GAG AGG CTA GCC AGC TGC CGA CCC CTT ACC GAT TTT GAC CAG     385
Cys Pro Glu Arg Leu Ala Ser Cys Arg Pro Leu Thr Asp Phe Asp Gln
        115                 120                 125
GGC TGG GGC CCT ATC AGT TAT GCC AAC GGA AGC GGC CCC GAC CAG CGC     433
```

```
Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Pro Asp Gln Arg
    130                 135                 140
CCC TAC TGC TGG CAC TAC CCC CCA AAA CCT TGC GGT ATT GTG CCC GCG     481
Pro Tyr Cys Trp His Tyr Pro Pro Lys Pro Cys Gly Ile Val Pro Ala
145                 150                 155                 160
AAG AGT GTG TGT GGT CCG GTA TAT TGC TTC ACT CCC AGC CCC GTG GTG     529
Lys Ser Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
                165                 170                 175
GTG GGA ACG ACC GAC AGG TCG GGC GCG CCC ACC TAC AGC TGG GGT GAA     577
Val Gly Thr Thr Asp Arg Ser Gly Ala Pro Thr Tyr Ser Trp Gly Glu
            180                 185                 190
AAT GAT ACG GAC GTC TTC GTC CTT AAC AAT ACC AGG CCA CCG CTG GGC     625
Asn Asp Thr Asp Val Phe Val Leu Asn Asn Thr Arg Pro Pro Leu Gly
        195                 200                 205
AAT TGG TTC GGT TGT ACC TGG ATG AAC TCA ACT GGA TTC ACC AAA GTG     673
Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Val
    210                 215                 220
TGC GGA GCG CCT CCT TGT GTC ATC GGA GGG GCG GGC AAC AAC ACC CTG     721
Cys Gly Ala Pro Pro Cys Val Ile Gly Gly Ala Gly Asn Asn Thr Leu
225                 230                 235                 240
CAC TGC CCC ACT GAT TGC TTC CGC AAG CAT CCG GAC GCC ACA TAC TCT     769
His Cys Pro Thr Asp Cys Phe Arg Lys His Pro Asp Ala Thr Tyr Ser
                245                 250                 255
CGG TGC GGC TCC GGT CCC TGG ATC ACA CCC AGG TGC CTG GTC GAC TAC     817
Arg Cys Gly Ser Gly Pro Trp Ile Thr Pro Arg Cys Leu Val Asp Tyr
            260                 265                 270
CCG TAT AGG CTT TGG CAT TAT CCT TGT ACC ATC AAC TAC ACC ATA TTT     865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Ile Asn Tyr Thr Ile Phe
        275                 280                 285
```

```
AAA ATC AGG ATG TAC GTG GGA GGG GTC GAA CAC AGG CTG GAA GCT GCC     913
Lys Ile Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Glu Ala Ala
    290                 295                 300
TGC AAC TGG ACG CGG GGC GAA CGT TGC GAT CTG GAA GAC AGG GAC AGG     961
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320
TCC GAG CTC AGC CCG TTA CTG CTG ACC ACT ACA CAG TGG CAG GTC CTC    1009
Ser Glu Leu Ser Pro Leu Leu Leu Thr Thr Thr Gln Trp Gln Val Leu
                325                 330                 335
CCG TGT TCC TTC ACA ACC CTA CCA GCC TTG TCC ACC GGC CTC ATC CAC    1057
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350
CTC CAC CAG AAC ATT GTG GAC GTG CAG TAC TTG TAC GGG GTG GGG TCA    1105
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Val Gly Ser
        355                 360                 365
AGC ATC GCG TCC TGG GCC ATT AAG TGG GAG TAC GTC GTT CTC CTG TTC    1153
Ser Ile Ala Ser Trp Ala Ile Lys Trp Glu Tyr Val Val Leu Leu Phe
    370                 375                 380
CTT CTG CTT GCA GAC GCG CGC GTC TGC TCC TGC TTG TGG ATG ATG CTA    1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ser Cys Leu Trp Met Met Leu
385                 390                 395                 400
CTC ATA                                                            1207
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: H77

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
G ATC CCA CAA GCC ATC ATG GAC ATG ATC GCT GGT GCT CAC TGG GGA GTC     49
  Ile Pro Gln Ala Ile Met Asp Met Ile Ala Gly Ala His Trp Gly Val
    1               5                  10                  15

CTG GCG GGC ATA GCG TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC      97
Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30

CTG GTA GTG CTG CTG CTA TTT GCC GGC GTC GAC GCG GAA ACC CAC GTC     145
Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala Glu Thr His Val
        35                  40                  45

ACC GGG GGA AGT GCC GGC CGC ACC ACG GCT GGG CTT GTT GGT CTC CTT     193
Thr Gly Gly Ser Ala Gly Arg Thr Thr Ala Gly Leu Val Gly Leu Leu
    50                  55                  60

ACA CCA GGC GCC AAG CAG AAC ATC CAA CTG ATC AAC ACC AAC GGC AGT     241
```

```
Thr Pro Gly Ala Lys Gln Asn Ile Gln Leu Ile Asn Thr Asn Gly Ser
 65                  70                  75                  80
TGG CAC ATC AAT AGC ACG GCC TTG AAC TGC AAT GAA AGC CTT AAC ACC     289
Trp His Ile Asn Ser Thr Ala Leu Asn Cys Asn Glu Ser Leu Asn Thr
                85                  90                  95
GGC TGG TTA GCA GGG CTC TTC TAT CAC CAC AAA TTC AAC TCT TCA GGC     337
Gly Trp Leu Ala Gly Leu Phe Tyr His His Lys Phe Asn Ser Ser Gly
            100                 105                 110
TGT CCT GAG AGG TTG GCC AGC TGC CGA CGC CTT ACC GAT TTT GCC CAG     385
Cys Pro Glu Arg Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln
        115                 120                 125
GGC TGG GGT CCT ATC AGT TAT GCC AAC GGA AGC GGC CTC GAC GAA CGC     433
Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Leu Asp Glu Arg
    130                 135                 140
CCC TAC TGC TGG CAC TAC CCT CCA AGA CCT TGT GGC ATT GTG CCC GCA     481
Pro Tyr Cys Trp His Tyr Pro Pro Arg Pro Cys Gly Ile Val Pro Ala
145                 150                 155                 160
AAG AGC GTG TGT GGC CCG GTA TAT TGC TTC ACT CCC AGC CCC GTG GTG     529
Lys Ser Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
                165                 170                 175
GTG GGA ACG ACC GAC AGG TCG GGC GCG CCT ACC TAC AGC TGG GGT GCA     577
Val Gly Thr Thr Asp Arg Ser Gly Ala Pro Thr Tyr Ser Trp Gly Ala
            180                 185                 190
AAT GAT ACG GAT GTC TTC GTC CTT AAC AAC ACC AGG CCA CCG CTG GGC     625
Asn Asp Thr Asp Val Phe Val Leu Asn Asn Thr Arg Pro Pro Leu Gly
        195                 200                 205
AAT TGG TTC GGT TGT ACC TGG ATG AAC TCA ACT GGA TTC ACC AAA GTG     673
Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Val
    210                 215                 220
```

```
TGC GGA GCG CCC CCT TGT GTC ATC GGA GGG GTG GGC AAC AAC ACC TTG     721
Cys Gly Ala Pro Pro Cys Val Ile Gly Gly Val Gly Asn Asn Thr Leu
225                 230                 235                 240
CTC TGC CCC ACT GAT TGC TTC CGC AAG CAT CCG GAA GCC ACA TAC TCT     769
Leu Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Ser
                245                 250                 255
CGG TGC GGC TCC GGT CCC TGG ATT ACA CCC AGG TGC ATG GTC GAC TAC     817
Arg Cys Gly Ser Gly Pro Trp Ile Thr Pro Arg Cys Met Val Asp Tyr
            260                 265                 270
CCG TAT AGG CTT TGG CAC TAT CCT TGT ACC ATC AAT TAC ACC ATA TTC     865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Ile Asn Tyr Thr Ile Phe
        275                 280                 285
AAA GTC AGG ATG TAC GTG GGA GGG GTC GAG CAC AGG CTG GAA GCG GCC     913
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Glu Ala Ala
    290                 295                 300
TGC AAC TGG ACG CGG GGC GAA CGC TGT GAT CTG GAA GAC AGG GAC AGG     961
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320
TCC GAG CTC AGC CCA TTG CTG CTG TCC ACC ACA CAG TGG CAG GTC CTT    1009
Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Gln Trp Gln Val Leu
                325                 330                 335
CCG TGT TCT TTC ACG ACC CTG CCA GCC TTG TCC ACC GGC CTC ATC CAC    1057
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350
CTC CAC CAG AAC ATT GTG GAC GTG CAG TAC TTG TAC GGG GTA GGG TCA    1105
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Val Gly Ser
        355                 360                 365
AGC ATC GCG TCC TGG GCC ATT AAG TGG GAG TAC GTC GTT CTC CTG TTC    1153
Ser Ile Ala Ser Trp Ala Ile Lys Trp Glu Tyr Val Val Leu Leu Phe
```

```
    370                 375                 380
CTT CTG CTT GCA GAC GCG CGC GTC TGC TCC TGC TTG TGG ATG ATG TTA    1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ser Cys Leu Trp Met Met Leu
385                 390                 395                 400
CTC ATA                                                        1207
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: H90

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
G ATC CCA CAA GCC ATC ATG GAT ATG ATC GCT GGT GCT CAC TGG GGA GTC    49
  Ile Pro Gln Ala Ile Met Asp Met Ile Ala Gly Ala His Trp Gly Val
    1               5                  10                  15
CTG GCG GGC ATA GCG TAT TTC TCC ATG GTA GGG AAC TGG GCG AAG GTC      97
Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30
```

```
CTA GTA GTG CTG CTG CTA TTT GCC GGC GTC GAC GCG GAA ACC CAC GTC     145
Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala Glu Thr His Val
         35                  40                  45
ACC GGG GGA AGT GCC GGC CGC TCC GTG CTT GGG ATT GCT AGT TTC CTT     193
Thr Gly Gly Ser Ala Gly Arg Ser Val Leu Gly Ile Ala Ser Phe Leu
     50                  55                  60
ACA CGA GGC CCC AAG CAG AAC ATC CAG CTG ATC AAA ACC AAC GGC AGT     241
Thr Arg Gly Pro Lys Gln Asn Ile Gln Leu Ile Lys Thr Asn Gly Ser
 65                  70                  75                  80
TGG CAC ATC AAT AGC ACG GCC CTG AAC TGC AAT GAC AGC CTT AAC GCC     289
Trp His Ile Asn Ser Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn Ala
                 85                  90                  95
GGC TGG ATA GCG GGG CTC TTC TAT CAC CAT GGA TTC AAC TCT TCA GGC     337
Gly Trp Ile Ala Gly Leu Phe Tyr His His Gly Phe Asn Ser Ser Gly
            100                 105                 110
TGT CCT GAG AGG TTG GCC AGC TGC CGA CGC CTT ACC GAT TTT GAC CAG     385
Cys Pro Glu Arg Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Asp Gln
        115                 120                 125
GGC TGG GGC CCT ATC AGT TAT GCC AAC GGA AGC GGC CCC GAC GAA CGT     433
Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Pro Asp Glu Arg
    130                 135                 140
CCC TAC TGC TGG CAC TAC CCC CCA AGA CCT TGT GGC ATT GTG CCC GCA     481
Pro Tyr Cys Trp His Tyr Pro Pro Arg Pro Cys Gly Ile Val Pro Ala
145                 150                 155                 160
AAG AGC GTG TGT GGC CCG GTA TAC TGC TTC ACT CCC AGC CCC GTG GTG     529
Lys Ser Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
                165                 170                 175
GTG GGA ACG ACC GAC AGG TCG GGC GCG CCT ACC TAC AAC TGG GGT GAA     577
Val Gly Thr Thr Asp Arg Ser Gly Ala Pro Thr Tyr Asn Trp Gly Glu
```

```
                180                 185                 190
AAT GAT ACG GAT GTC CTC ATC CTT AAC AAC ACC AGG CCG CCG CTG GGC     625
Asn Asp Thr Asp Val Leu Ile Leu Asn Asn Thr Arg Pro Pro Leu Gly
        195                 200                 205
AAT TGG TTC GGT TGT ACC TGG ATG AAC TCA ACT GGA TTC ACC AAA GTG     673
Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Val
    210                 215                 220
TGC GGA GCG CCC CCT TGT GTC ATC GGA GGG GTG GGC AAC AAC ACC TTG     721
Cys Gly Ala Pro Pro Cys Val Ile Gly Gly Val Gly Asn Asn Thr Leu
225                 230                 235                 240
CGC TGC CCC ACT GAT TGT TTC CGC AAG CAT CCG GAA GCC ACA TAC TCT     769
Arg Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Ser
                245                 250                 255
CGG TGC GGC TCC GGT CCC TGG ATC ACA CCC AGG TGC ATG GTC CAC TAC     817
Arg Cys Gly Ser Gly Pro Trp Ile Thr Pro Arg Cys Met Val His Tyr
            260                 265                 270
CCG TAT AGG CTT TGG CAC TAT CCT TGT ACC ATC AAT TAC ACT ATA TTT     865
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Ile Asn Tyr Thr Ile Phe
        275                 280                 285
AAA GTC AGG ATG TAC GTG GGA GGG ATC GAG CAC AGG CTG GAA GCG GCC     913
Lys Val Arg Met Tyr Val Gly Gly Ile Glu His Arg Leu Glu Ala Ala
    290                 295                 300
TGC AAC TGG ACG CGG GGC GAA CGT TGC GAT CTG GAA GAC AGG GAC AGG     961
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
305                 310                 315                 320
TCC GAG CTC AGC CCA TTG CTG CTG TCC ACT ACG CAG TGG CAG GTC CTT    1009
Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Gln Trp Gln Val Leu
                325                 330                 335
CCG TGT TCT TTC ACG ACC CTG CCA GCC TTG TCC ACC GGC CTC ATC CAC    1057
```

```
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
            340                 345                 350

CTC CAC CAG AAC ATT GTG GAC GTG CAG TAC TTG TAC GGG GTA GGG TCA    1105
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Val Gly Ser
        355                 360                 365

AGC ATC GCG TCC TGG ACC ATC AAG TGG GAG TAC GTC GTT CTC CTG TTC    1153
Ser Ile Ala Ser Trp Thr Ile Lys Trp Glu Tyr Val Val Leu Leu Phe
    370                 375                 380

CTC CTG CTT GCA GAC GCG CGC GTC TGC TCC TGC TTG TGG ATG ATG TTA    1201
Leu Leu Leu Ala Asp Ala Arg Val Cys Ser Cys Leu Trp Met Met Leu
385                 390                 395                 400

CTC ATA                                                            1207
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 523 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: J1(JM)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
G ATC CCA CAA GCC ATC TTG GAT ATG ATC GCT GGT GCT CAC TGG GGA GTC     49
  Ile Pro Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val
    1               5                  10                  15
CTG GCG GGC ATA GCG TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC      97
Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30
CTG GTA GTG CTG TTG CTG TTT GCC GGC GTC GAC GCG GAA ACC ATC GTC     145
Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala Glu Thr Ile Val
        35                  40                  45
TCC GGG GGA CAA GCC GCC CGC GCC ATG TCT GGA CTT GTT AGT CTC TTC     193
Ser Gly Gly Gln Ala Ala Arg Ala Met Ser Gly Leu Val Ser Leu Phe
    50                  55                  60
ACA CCA GGC GCT AAG CAG AAC ATC CAG CTG ATC AAC ACC AAC GGC AGT     241
Thr Pro Gly Ala Lys Gln Asn Ile Gln Leu Ile Asn Thr Asn Gly Ser
65                  70                  75                  80
TGG CAC ATC AAT AGC ACG GCC TTG AAC TGC AAT GAA AGC CTT AAC ACC     289
Trp His Ile Asn Ser Thr Ala Leu Asn Cys Asn Glu Ser Leu Asn Thr
                85                  90                  95
GGC TGG TTA GCA GGG CTT ATC TAT CAA CAC AAA TTC AAC TCT TCG GGC     337
Gly Trp Leu Ala Gly Leu Ile Tyr Gln His Lys Phe Asn Ser Ser Gly
            100                 105                 110
TGT CCC GAG AGG TTG GCC AGC TGC CGA CGC CTT ACC GAT TTT GAC CAG     385
Cys Pro Glu Arg Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Asp Gln
        115                 120                 125
GGC TGG GGC CCT ATC AGT CAT GCC AAC GGA AGC GGC CCC GAC CAA CGC     433
Gly Trp Gly Pro Ile Ser His Ala Asn Gly Ser Gly Pro Asp Gln Arg
    130                 135                 140
CCC TAT TGT TGG CAC TAC CCC CCA AAA CCT TGC GGT ATC GTG CCC GCA     481
```

```
Pro Tyr Cys Trp His Tyr Pro Pro Lys Pro Cys Gly Il  Val Pro Ala
145                 150                 155                 160

AAG AGC GTA TGT GGC CCG GTA TAT TGC TTC ACT CCC AGC CCC          523
Lys Ser Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro
                165                 170
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 523 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(B) CLONE: J4(JM)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
G ATC CCA CAA GCT GTC GTG GAC ATG GTG GCG GGG GCC CAC TGG GGA GTC    49
  Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val
    1               5                  10                  15

CTG GCG GGC CTT GCC TAC TAT TCC ATG GTA GGG AAC TGG GCT AAG GTC    97
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
            20                  25                  30

CTG ATT GTG GCG CTA CTC TTC GCC GGC GTT GAC GGG GAG ACC TAC ACG    145
```

```
Leu Ile Val Ala Leu Leu Phe Ala Gly Val Asp Gly Glu Thr Tyr Thr
        35                  40                  45

TCG GGG GGG GCG GCC AGC CAC ACC ACC TCC ACG CTC GCG TCC CTC TTC     193
Ser Gly Gly Ala Ala Ser His Thr Thr Ser Thr Leu Ala Ser Leu Phe
    50                  55                  60

TCA CCT GGG GCG TCT CAG AGA ATC CAG CTT GTG AAT ACC AAC GGC AGC     241
Ser Pro Gly Ala Ser Gln Arg Ile Gln Leu Val Asn Thr Asn Gly Ser
65                  70                  75                  80

TGG CAC ATC AAC AGG ACT GCC CTA AAC TGC AAT GAC TCC CTC CAC ACT     289
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu His Thr
                85                  90                  95

GGG TTC CTT GCC GCG CTG TTC TAC ACA CAC AGG TTC AAC TCG TCC GGG     337
Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Arg Phe Asn Ser Ser Gly
            100                 105                 110

TGC CCG GAG CGC ATG GCC AGC TGC CGC CCC ATT GAC TGG TTC GCC CAG     385
Cys Pro Glu Arg Met Ala Ser Cys Arg Pro Ile Asp Trp Phe Ala Gln
        115                 120                 125

GGA TGG GGC CCC ATC ACC TAT ACT GAG CCT GAC AGC CCG GAT CAG AGG     433
Gly Trp Gly Pro Ile Thr Tyr Thr Glu Pro Asp Ser Pro Asp Gln Arg
    130                 135                 140

CCT TAT TGC TGG CAT TAC GCG CCT CGA CCG TGT GGT ATC GTA CCC GCG     481
Pro Tyr Cys Trp His Tyr Ala Pro Arg Pro Cys Gly Ile Val Pro Ala
145                 150                 155                 160

TCG CAG GTG TGT GGT CCA GTG TAT TGC TTC ACC CCA AGC CCT             523
Ser Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro
                165                 170
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 402 amino acids

(B) TYPE: protein

(vi) ORIGINAL SOURCE:

(A) ORIGIN: Hepatitis C virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val
              Ile Met         Ile
                  Leu
 1               5                  10                  15
Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
           Ile         Phe         Ala
            20                  25                  30
Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His Val
   Ile     Ala                             Ala Arg     Arg Thr
           Leu                                 His     Tyr Arg
                                               Asp     Ile
                                               Gln     Gln
                                               Glu     His
                                               Thr     Thr
        35                  40                  45
Thr Gly Gly Lys Val Ala Tyr Thr Thr Gln Gly Phe Thr Pro Phe Phe
Ser Val Ala Val Gln Gly His Val Val Ser Arg Leu Val Ser Leu Leu
Met         Ala Ala Ser Lys Ser Met Asn Ser Val Ala Arg Met
            Arg         Ser Ala Ala Thr Thr Ile     Gly Trp
            Ser         Arg Arg     His Ala
```

```
                Gln          Phe Gly     Leu
                His              His     Tyr
                Asn                      Ala
                Ile

     50                  55                  60
Ser Arg Gly Pro Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly Ser
Arg Pro     Ala Gln Glu Arg Val         Ile Lys
Thr Ser     Ser Lys     Asn
Ala Gln         Ala     Asp
Asn Ala         Arg
    Leu
 65                  70                  75                  80
Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn Thr
        Leu     Ser                         Glu         Gln Ala
                                                        Lys
                                                        His
                  85                  90                  95
Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Ser Phe Asn Ala Ser Gly
    Trp Ile     Thr     Ile     Arg     Lys     Asp Ser     Arg
                Gly             Ala     Arg
                                His     Gly
                                Gln
            100                 105                 110
Cys Pro Glu Arg Met Ala Gly Cys Arg Pro Ile Asp Glu Phe Ala Gln
Ala Glu Ser Val Leu     Ser     Cys Ser Leu Ser Lys     Asp
                        Gln     Gln Thr     Thr Trp
                                    Phe         Asp
                                    Arg         Thr
        115                 120                 125
```

```
Gly Trp Gly Pro Ile Thr His Val Val Pro Asn Ile Ser Asp Gln Arg
        Asp         Ser Tyr Ala Gln Ser Asp Val Pro Glu Glu Lys
                            Asp Glu Arg Ser Asn Thr
                            Thr Met Gly Glu Arg Gly
                            Asn Asn Gln Arg Ser
                                Lys     Gly Gly
                                        Thr
    130                 135                 140
Pro Tyr Cys Trp His Tyr Ala Pro Arg Pro Cys Gly Ile Val Pro Ala
                        Pro     Pro Gln     Thr Val
                                Lys
145                 150                 155                 160
Ser Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val
Trp Glu
Lys Ser
                165                 170                 175
Val Gly Thr Thr Asp Arg Phe Gly Ala Pro Thr Tyr Asn Trp Gly Asn
                        Ser     Val             Thr         Ala
                                                Arg         Asp
                                                Ser         Glu
            180                 185                 190
Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr Arg Pro Pro Gln Gly
    Asp             Phe Val     Ser                     His
                        Ile                             Leu
        195                 200                 205
Asn Trp Phe Gly Cys Thr Trp Met Asn Gly Thr Gly Phe Thr Lys Thr
                                    Ser                     Val
    210                 215                 220
Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Val Gly Asn Asn Thr Leu
```

```
        Ala             His     Arg     Arg
                        Val             Ala
225                 230                 235                 240
Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Thr
Val                                         Asp             Ser
His
Leu
Arg
                245                 250                 255
Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu Val His Tyr
Arg                         Ile                 Met     Asp
            260                 265                 270
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Ile Phe
                                        Ile     Tyr     Val
                                                        Thr
        275                 280                 285
Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Glu Ala Ala
    Ile                         Ile                 Asp
                                                    Asn
                                                    Ile
    290                 295                 300
Cys Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg
                        Asp         Asn
305                 310                 315                 320
Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp Gln Val Leu
Ala                             Thr         Gln         Ile
Pro
                325                 330                 335
Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His
```

```
                340                 345                 350
Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser
        Arg     Val                                     Val
        355                 360                 365
Ala Val Val Ser Phe Ala Ile Lys Trp Glu Tyr Ile Leu Leu Leu Phe
Ser Ile Ala     Ile Val     Arg             Val Val
                Trp Thr
    370                 375                 380
Leu Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu
                                        Ser
385                 390                 395                 400
Leu Ile
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: the other nucleic acid (synthesized DNA for PCR)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCTATCAGCA GCATCATCCA 20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: the other nucleic acid (synthesized DNA for PCR)

( ) SEQUENCE CHARACTERISTIC: N represents inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CAGNTANTCC GGATCCCNCA AG 22

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: the other nucleic acid (synthesized DNA for PCR)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTAAAACGAC GGCCAGT 17

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: the other nucleic acid (synthesized DNA for PCR)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAGGAAACAG CTATGAC 17

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: the other nucleic acid (synthesized DNA for PCR)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGACTAGTCC 10

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: the other nucleic acid (synthesized DNA for PCR)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTAGAGAATT CGGTAC 16

Claims

1. A diagnostic reagent for hepatitis C, which detects an antibody induced by infection of hepatitis C virus, characterized in that said diagnostic reagent comprises the second envelope protein or the first non-structural protein which is encoded by the gene of hepatitis C virus and has a suger chain.

2. The diagnostic reagent for hepatitis C according to Claim 1, wherein the second envelope protein or the first non-structural protein is represented by an amino acid sequence selected from SEQUENCE ID Nos.1-12 of SEQUENCE LISTING or an amino acid sequence having the same as a part thereof.

3. The diagnostic reagent for hepatitis C according to Claim 1, wherein the second envelope protein or the first non-structural protein is encoded by a base sequence selected from SEQUENCE ID Nos.1-2, and 4-11 of SEQUENCE LISTING.

4. The diagnostic reagent for hepatitis C according to Claim 1, wherein the second envelope protein or the first non-structural protein is produced by an animal cell.

5. The diagnostic reagent for hepatitis C according to Claim 4, wherein the animal cell is CHO cell.

6. A method for detecting an anti-hepatitis C virus antibody, wherein the second envelope protein or the first non-structural protein which is encoded by th gene of hepatitis C virus and has a suger chain is used as an antigen to detect the antibody specific to said antigen.

7. The method according to Claim 6, wherein the s cond envelop protein or the first non-structural protein is repr sented by an amino acid s quenc s l cted from SEQUENCE ID Nos.1-12 of SEQUENCE LISTING or an amino acid sequence having the same as a part thereof.

8. The method according to Claim 6, wherein the second envelope protein or the first non-structural protein is encoded by a base sequence selected from SEQUENCE ID Nos.1-2, and 4-11 of SEQUENCE LISTING.

9. The method according to Claim 9, wherein the second envelope protein or the first non-structural protein is produced by an animal cell.

10. The method according to Claim 6, wherein the animal cell is CHO cell.

11. A method for detecting an anti-hepatitis C virus antibody, which comprises the steps of contacting a sample with the second envelope protein or the first non-structural protein which is encoded by the gene of hepatitis C virus and has a suger chain under the conditions that the second envelope protein or the first non-structural protein is bound to the anti-hepatitis C virus antibody to form an immunological complex and measuring the formation of the immunological complex to confirm the presence of the anti-hepatitis C virus antibody in the sample.

12. The method according to Claim 11, wherein the formation of the immunological complex is measured by RIA, ELISA, fluorescent antibody technique, agglutination reaction, or immune precipitation.

FIG. 1

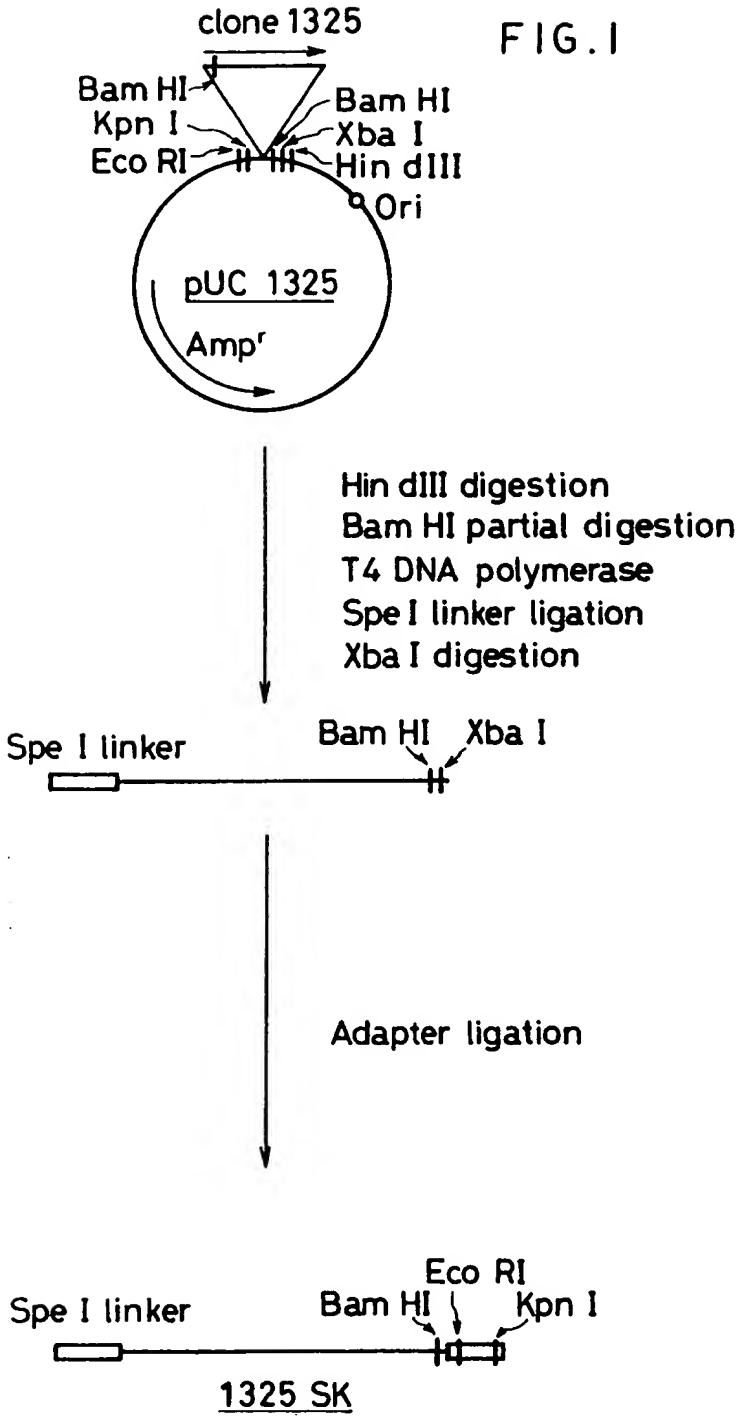

clone 1325
Bam HI
Kpn I
Eco RI
Bam HI
Xba I
Hin dIII
Ori
pUC 1325
Amp^r
Hin dIII digestion
Bam HI partial digestion
T4 DNA polymerase
Spe I linker ligation
Xba I digestion
Spe I linker
Bam HI
Xba I
Adapter ligation
Eco RI
Spe I linker
Bam HI
Kpn I
1325 SK

FIG.2

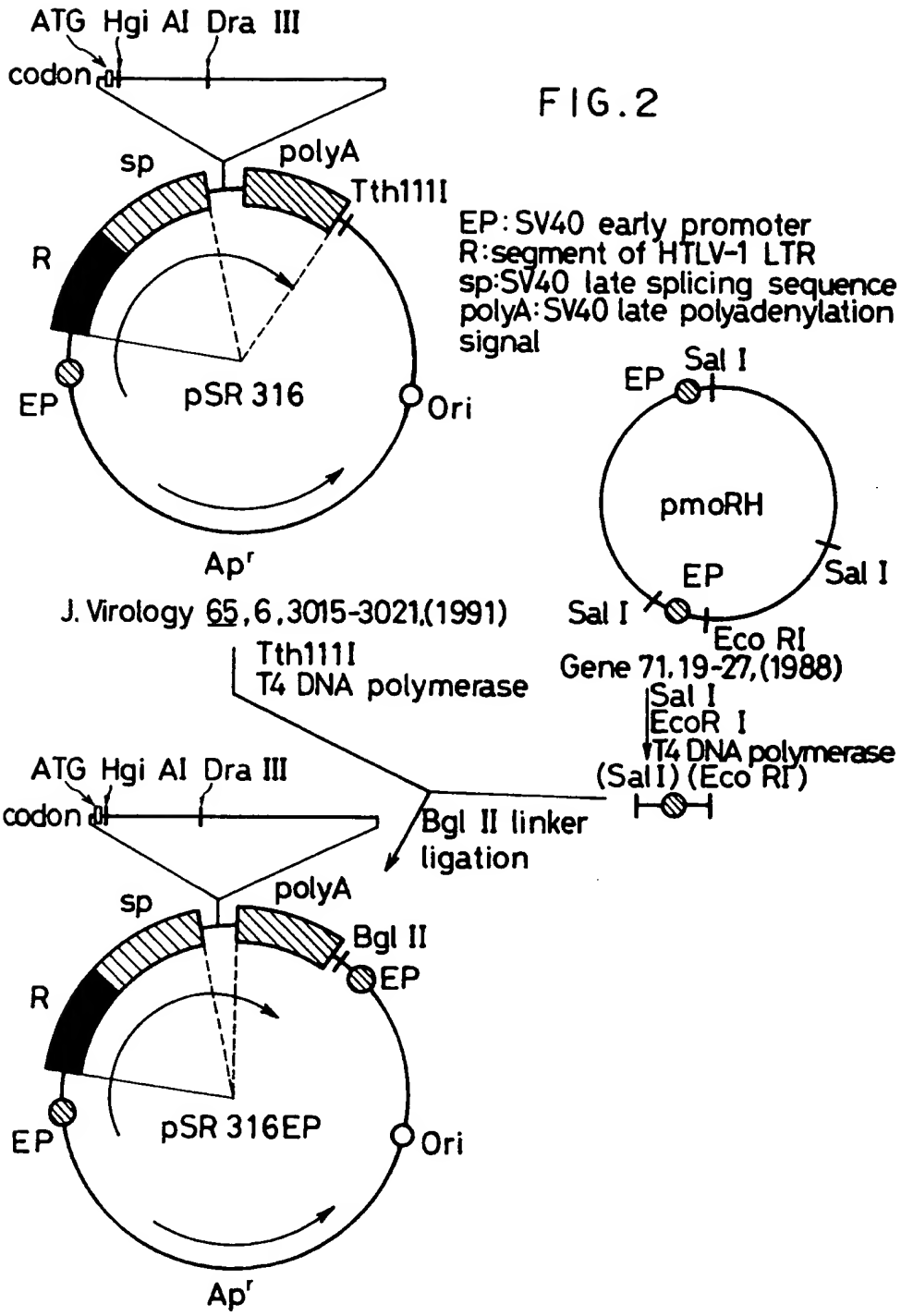

ATG Hgi AI Dra III
codon
sp
polyA
Tth111I
R
EP
pSR 316
Ori
Ap^r
EP: SV40 early promoter
R: segment of HTLV-1 LTR
sp: SV40 late splicing sequence
polyA: SV40 late polyadenylation signal
J. Virology 65, 6, 3015-3021, (1991)
Tth111I
T4 DNA polymerase
Sal I
EP
pmoRH
EP
Sal I
Sal I
Eco RI
Gene 71, 19-27, (1988)
Sal I
EcoR I
T4 DNA polymerase
(Sal I) (Eco RI)
Bgl II linker
ligation
ATG Hgi AI Dra III
codon
sp
polyA
Bgl II
EP
R
EP
pSR 316EP
Ori
Ap^r

FIG.3

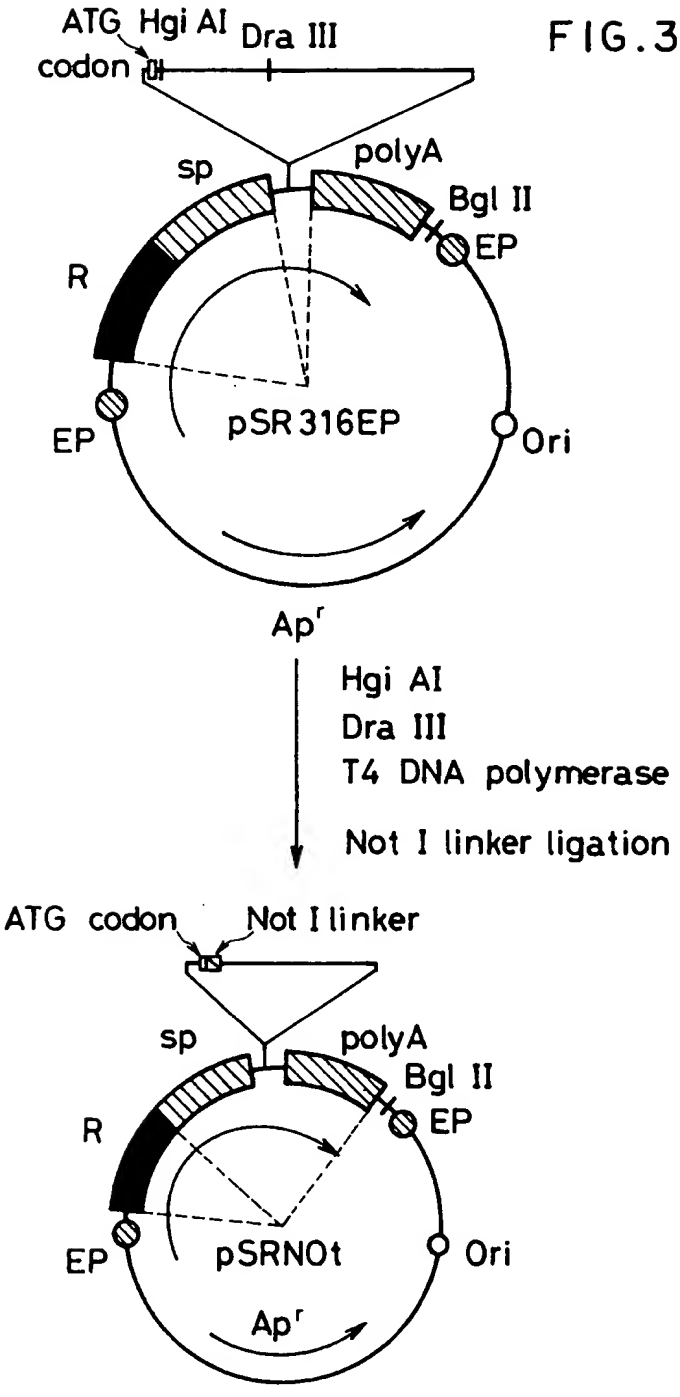

ATG codon
Hgi AI
Dra III
sp
polyA
Bgl II
EP
R
EP
pSR316EP
Ori
Ap^r
Hgi AI
Dra III
T4 DNA polymerase
Not I linker ligation
ATG codon
Not I linker
sp
polyA
Bgl II
EP
R
EP
pSRNOt
Ori
Ap^r

FIG. 4

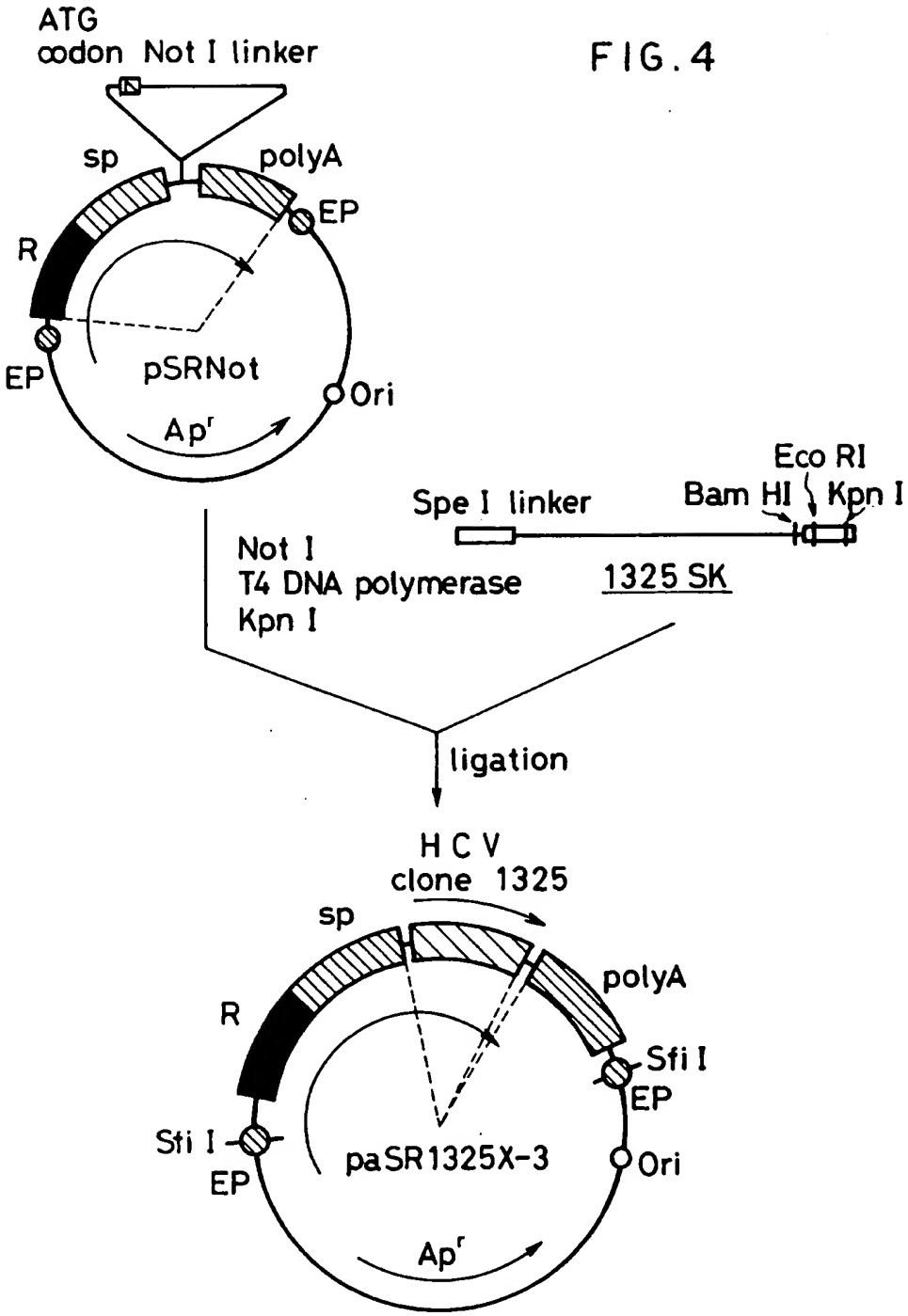

ATG
codon Not I linker
sp
polyA
EP
R
EP
pSRNot
Ori
Ap^r
Eco RI
Spe I linker
Bam HI
Kpn I
Not I
T4 DNA polymerase
Kpn I
1325 SK
ligation
H C V
clone 1325
sp
polyA
R
Sfi I
EP
Sfi I
EP
paSR1325X-3
Ori
Ap^r

FIG.5

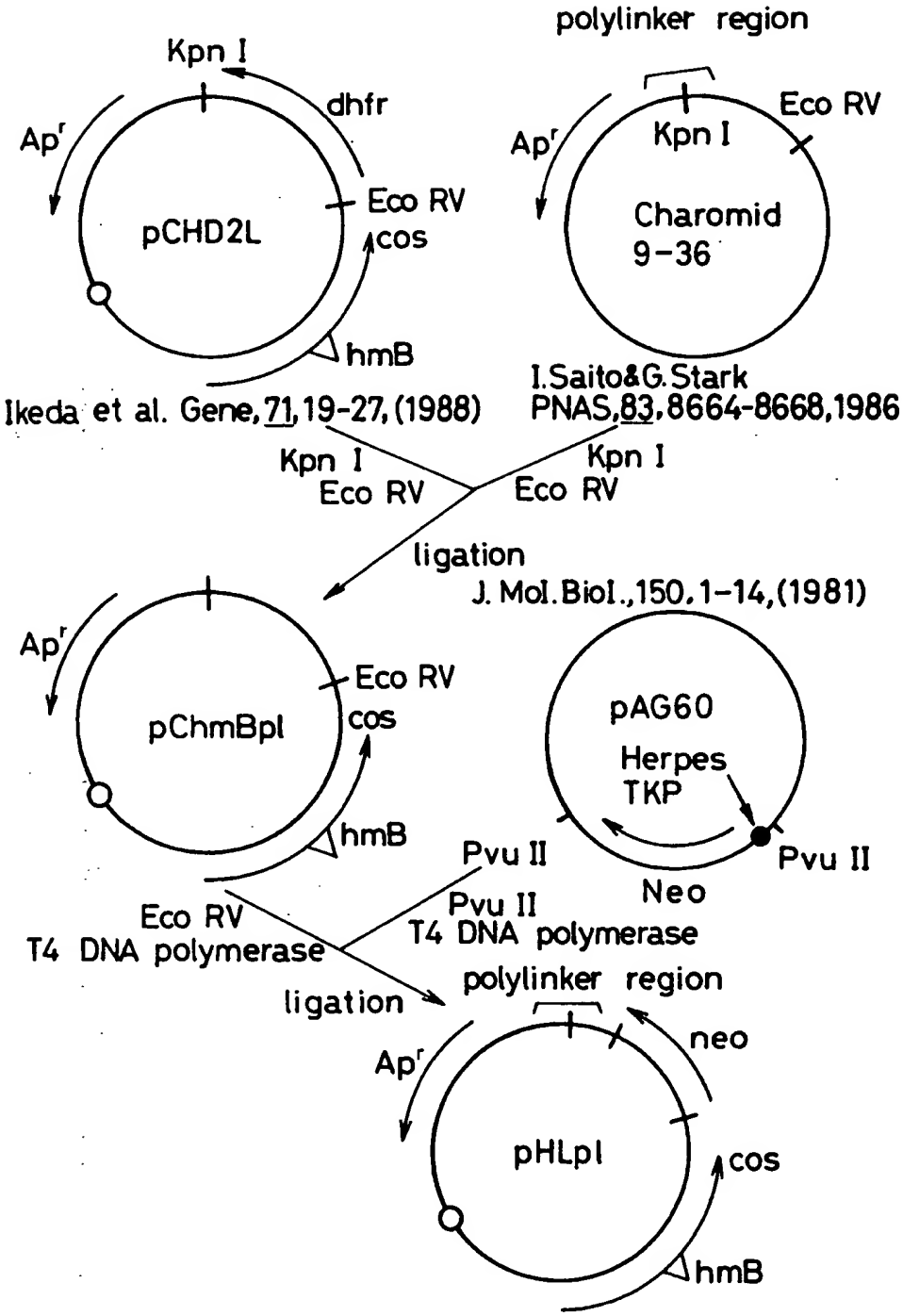
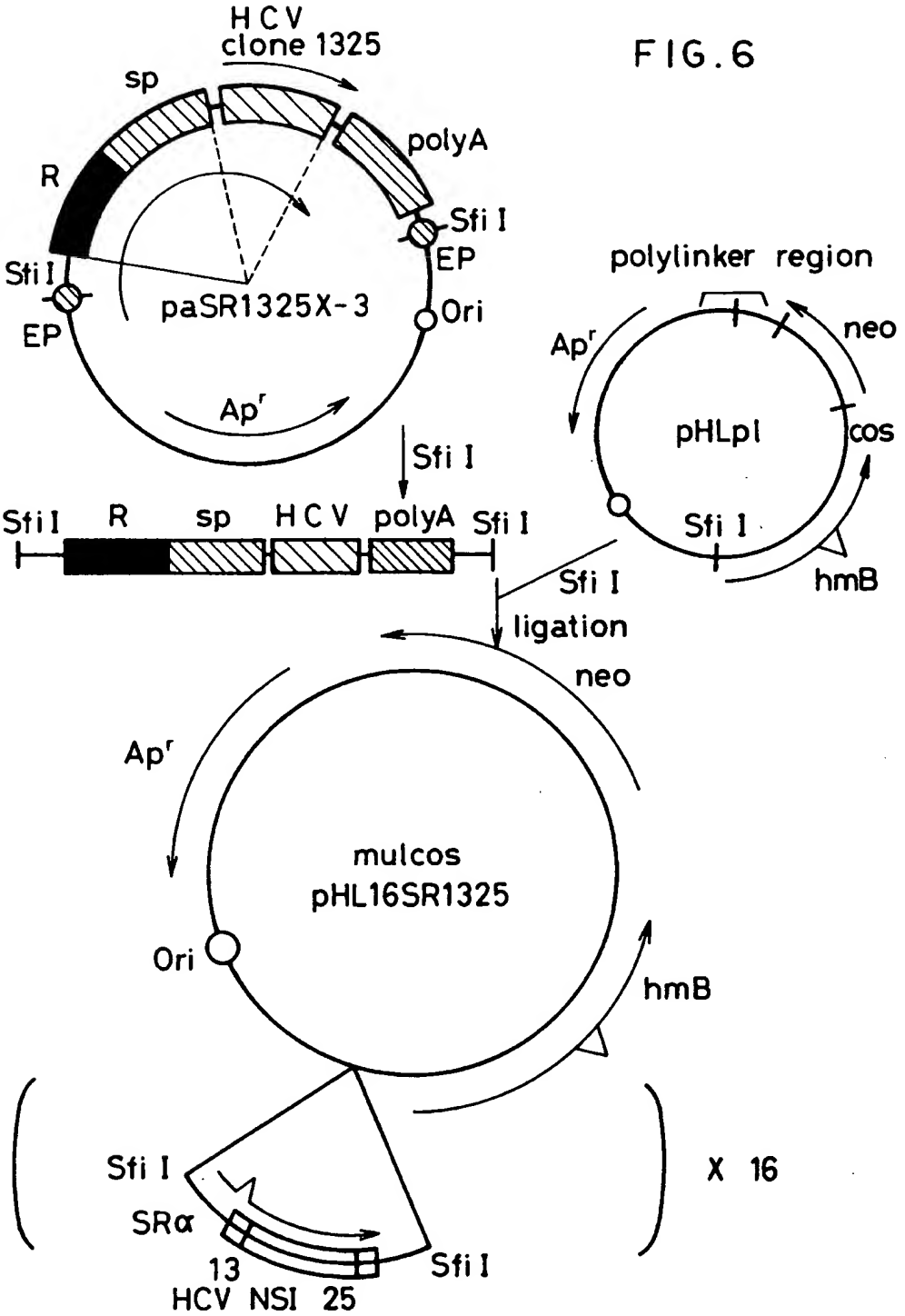

polylinker region
Kpn I
dhfr
Ap^r
Eco RV
pCHD2L
cos
hmB
Ikeda et al. Gene, 71, 19-27, (1988)
Eco RV
Kpn I
Charomid
9-36
I.Saito&G.Stark
PNAS, 83, 8664-8668, 1986
Kpn I
Eco RV
ligation
J. Mol. Biol., 150, 1-14, (1981)
pChmBpl
pAG60
Herpes
TKP
Pvu II
Neo
Eco RV
T4 DNA polymerase
Pvu II
T4 DNA polymerase
ligation
polylinker region
neo
pHLpl

FIG.6


HCV
clone 1325
sp
polyA
R
Sfi I
EP
Sfi I
paSR1325X-3
EP
Ori
Ap^r
Sfi I
Sfi I R sp HCV polyA Sfi I
polylinker region
Ap^r
neo
pHLp1
cos
Sfi I
hmB
Sfi I
ligation
neo
Ap^r
mulcos
pHL16SR1325
Ori
hmB
Sfi I
SRα
13
HCV NSI 25
Sfi I
X 16

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 92 11 7191

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
X	F. BLAINE HOLLINGER 'VIRAL HEPATITIS AND LIVER DISEASE' 1 June 1991 , WILLIAMS & WILKENS , BALTIMORE MD USA See table I in article by G. Kuo et al. : "Serodiagnosis of hepatitis C viral infection using recombinant-based assays for circulating antibodies to different viral proteins." on page 347 - 349. ---	1	G01N33/576 C07K15/00
Y	EP-A-0 388 232 (CHIRON CORPORATION) * page 21, line 9 - line 27 * * page 34, line 6 - line 12 * ---	1-12	
Y	VIROLOGY vol. 180, 1 February 1991, WASHINGTON DC USA pages 842 - 848 A.J. WEINER ET AL. 'Variable and hypervariable domains are found in the regions of HCV corresponding to the Flavivirus envelope and NS1 proteins and the Pestivirus envelope glycoproteins.' * the whole document * ---	1-12	TECHNICAL FIELDS SEARCHED (Int. Cl.5) G01N C07K
T	HEPATOLOGY vol. 16, no. 4, 1992, WASHINGTON DC USA page 226A O. YOKOSUKA ET AL. 'Detection of anti-hepatitis c virus E2/NS1 antibody in patients with type c liver disease by western blotting.' * the whole document * -----	1-12	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	19 JANUARY 1993	VAN BOHEMEN C.G.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P0401)